



TAKING SURVEYS TO THE NEXT LEVEL:

When External Validation is a Good Thing

by [Robert A. Page, Jr., Ph.D.](#)

[Contact PDI](#) about these publications

(copyright 2002 by PDI. All rights reserved. Please do not modify, copy or distribute this article -- see website [terms and conditions of use.](#))

Most surveys are put together too quickly to realize their full potential in gathering and reporting useful, accurate and insightful information. This handout is designed to give you the information you need on the science of survey research to bring your surveys to the next level. Pragmatically, you will better understand the survey research process, what it offers, what its limitations are, the tradeoffs and pitfalls involved, and how to tell when survey findings are lying to you. Theoretically, you will be introduced to conceptual and statistical tools that can refine and improve the quality and validity of your survey questions and rating scales.

Survey Constructs: What's the Point?

Typically surveys promise to tell you about something you care about. The survey construct is supposed to describe a concept that is fairly abstract and complex - something you can not measure directly. Corporate culture and organizational effectiveness are two concepts managers care about, but are very difficult to measure. In fact, many researchers can not even agree on how they should be defined. This means that your opinion and perspective on such topics may be as good as anybody else's. A survey construct takes the concept you want to understand and operationalizes it, which is a fancy way of saying your concept becomes measurable. The categories, sub-categories and questions do the measuring with a rating scale. The individual questions should assess specific attitudes, activities and behaviors which describe their category. Ideally category questions should be similar enough in content so they make sense being grouped together, but not so similar that they are virtually identical, and tell you the same story.

Can I Trust What the Survey Tells Me?

"All models are bad, but some are useful"- A maxim in sociology

In terms of complete faith, you can never completely trust what a survey tells you. In survey research, the "Holy Grail" to search for is building "convergent validity." In its broadest and most rigorous sense, convergent validity means using information from a variety of sources to support each other, and "triangulate" on a research finding. When you have data from several different sources all pointing in the same direction, and telling you the same story, then you can have real confidence in your inferences and conclusions. Such data sources include:

- Surveys
- Written Comments (verbatim from open-ended questions)
- Accounting / Human Resources measures (productivity, absenteeism, etc.)
- Interviews and/or Focus groups
- Direct observation
- Archival data (Annual reports, company newspapers, minutes of meetings, etc.)

However, if you lack the time and money necessary to build compile comprehensive data from a variety of sources, there are sets of generally accepted survey research procedures and statistical tests that can help you decide how much confidence you should have in survey results alone.

What is the Sampling Strategy?

Survey results are only as good as the sample they came from. Ideally a researcher could use the whole population, and conduct a census where everyone in the organization would be asked to take the survey. However, given that population samples are expensive, time consuming, and will result in findings similar to those from a smaller, valid sample, there is seldom justification for population sampling - usually a good sample can do the job with a lot less hassle involved.

What is a valid sample? A valid sample accurately feeds back the viewpoints of the larger population it is supposed to represent. To increase the chance of this happening, survey researchers usually insist on a random sample, where everyone in the organization has an equal chance of being selected as a survey respondent. Further, given an expected response rate, the random sample must be sufficiently large to be statistically valid (95% confidence, 5% margin of error). So the first question to ask is simply: "Is the sample big enough?" Most polling companies use 1000 respondents as a default sample size for large organizations or communities.

Unfortunately, random samples are seldom appropriate for large, complex organizations. Most companies have a variety of groups who see things quite differently. Change agents usually recommend that all important groups be included in the sample so no relevant viewpoint is accidentally excluded. Statistically, this technique is called stratified random sampling. Each strata represents a different group, such as:

- Locations or Sites
- Functional Specializations
- Divisions or Departments
- Gender
- Ethnicity
- Length of Tenure
- Managerial Level or Position
- Business Units or Brands

So your next question becomes: "Are any groups excluded? Are any important perspectives on this topic being overlooked or excluded? What effect will their absence have on the findings?"

Unfortunately, there are no rules of thumb for calculating a stratified random sample - the process is complicated enough to require a statistical processing package. From a statistical standpoint, if you are planning to test statistical relationships and hypotheses later on, you need at least 35 respondents in the smallest cell of your demographic matrix for the calculations to work well. For example, if your demographics were gender, location and function, you would need 35 male and 35 female respondents from each function in each region.

Survey Scales

The most common interval scales are 5 point scales concerning agreement (how much does this statement characterize your work?) and frequency (how often does this happen?). The advantage of such scales are they are so general and all-purpose, you can ask a question concerning virtually any type of attitude or behavior without ever needing to change the rating scale:

1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree
 1 = Never, 2 = Seldom, 3 = Sometimes, 4 = Often, 5 = Always

Other interval scales are usually more specific and focused. Common examples include:

importance or priority scales (low versus high priority / importance)
 evaluation or comparison scales (poor versus excellent / above or below average)
 trend or improvement scales (better versus worse / great versus little)
 extent scales (to a greater versus lesser extent / degree)
 satisfaction scales (satisfied versus unsatisfied / attractive versus unattractive)

The danger with using a variety of scales in your survey instrument is that some respondents may either not notice that the rating scale has changed, or become confused. Clear transition paragraphs between scales help respondents keep track of the correct perceptual set (frame of reference). Careful data checking can identify respondents whose ratings did not change when the scale did.

There is also danger in not using a variety of scales in your survey instrument. Some academics argue that if you use a single rating scale, such as the agreement and frequency scales listed above, with positively worded items, respondents can stop discriminating between questions and simply answer all of them on the high end of the scale. Data checking must be diligent to eliminate this possibility, dropping respondents whose ratings feature large blocks of identical, uniform, and typically positive scores.

Likert scales combat this tendency by reversing the scale values, making low responses favorable and high responses unfavorable. For example, here are the two most common rating scales in a Likert format:

1 = Strongly Agree, 2 = Agree, 3 = Neutral, 4 = Disagree, 5 = Strongly Disagree
 1 = Always, 2 = Often, 3 = Sometimes, 4 = Seldom, 5 = Never

Critics suggest that this format is confusing, and is, in itself, a source of measurement error. The moral of this story is that there is no formula, and regardless of whatever rating scale strategy you choose, you must tread carefully, be aware of its weaknesses, and check your data for potential problems.

Even the best question can not overcome the handicap of a poor or inappropriate rating scale - the data become invalid. Common problems include:

Incompatibility. The phrasing of the scale does not match the phrasing of the question. This often happens when the scale and the questions are developed independently, and then linked together later, without careful editing.

Lack of Uniformity. The intervals between rating scale points must be uniform. For example, take a rating scale where 1= Never, 2=Occasionally, 3=Sometimes, and 4 = Always. The conceptual distance between rating points is not consistent, being greater on the ends of the scale and much smaller in the middle - points 2 and 3 are almost synonyms.

Lack of Symmetry. The favorable and unfavorable sides of a scale should be opposing reflections. There should be as many negative rating options as positive rating options, and the value labels should be antonyms (opposites). For example, take the scale:

1= Rarely, 2=Occasionally, 3=Sometimes, 4 = Frequently, 5=Always.

It has several symmetry problems: The opposite of 5 [Always] is not 1 [Rarely], it is "Never."
 The opposite of 4 [Frequently] is not 2 [Occasionally], it is 1 [Rarely].

Symmetry restored: 1= Never, 2= Rarely, 3= Sometimes, 4= Frequently, 5=Always, or
 1= Very Rarely, 2= Rarely, 3= Occasionally, 4= Frequently, 5= Very Frequently.

Inconsistency. The value labels given each rating point must measure the same type of rating, not different ones. For example, the following scale combines labels from a comparison scale, an effectiveness scale, and an evaluative scale:

1= Unacceptable, 2= Ineffective, 3= Average, 4= Effective, 5= Excellent

Interpreting the mean score becomes problematic, and the data is invalidated.

Ambiguity. The value labels given each rating point should be clear, specific, direct, and unambiguous. If they are vague, confusing, full of jargon or overly technical terms, the validity of the scale is compromised as respondents guess or assume meaning. For example, a rating scale anchored by "Very Good" versus "Very Bad" has ambiguity problems due to the highly subjective nature of the value judgments it is measuring - good and bad mean different things to different

respondents, and interpretations are unpredictable.

Interval Scale Length. 5 points is standard if respondents will likely make broad, general distinctions. 6 points is standard if respondents should have an opinion on the subject, and you do not give them the option of a neutral midpoint. 7 points is standard if respondents seem capable of making subtle distinctions.

Survey Construct Tradeoffs: Longer Surveys

Surveys will never give you all the information you need, because you tend to have far more questions than most respondents will want to answer. The longer the survey, the more information you will receive, and the more risks you take. These risks include:

Poor response rate - Respondents take a look and the length and refuse to answer. Low responses rates should be avoided, because they may have selection bias, where certain groups are over-represented. For example, if 30% of your respondents send the surveys back, and most of those respondents happen to be production people, this survey does not represent your organization, it represents one division or department within it.

Measurement error - Respondents become impatient or tired and get through the survey as quickly as possible, not putting much thought into their answers. They may give all the questions in entire categories the same score just to get it over with.

Negative attribution - Respondents come to regard surveys as more trouble than they are worth, and refuse to participate in future survey efforts.

Sometimes you need longer surveys because you need lots of information, and you need it now. There are strategies to avoid the pitfalls of poor response rates and poor quality data:

Build ownership. Make sure the people you target as respondents feel involved in the survey creation process, and they will be more likely to support the survey effort, even if the length is tiresome and a bit painful.

Consider incentives. Incentives can improve low response rates. At minimum, allow respondents to fill out the surveys on company time, or take compensatory time off. Incentives can range from vacation time to clothing to money to movie tickets, and do not have to be expensive. They can be individual (upon receiving your completed survey, we will send ...), group (when all team members return completed surveys, the team will receive ...) or inter-group (the first team to return completed surveys will receive ...)

Make commitments. If leadership commits to making changes on the basis of the survey results, respondents are likely to take it more seriously, and response rates should improve.

Track response rates. Follow-up letters, e-mail messages, company newspaper articles, and /or voice-mail messages can remind and encourage tardy respondents.

Survey Construct Tradeoffs: Shorter Surveys

While the obvious answer to this dilemma is to use short surveys, using short surveys offers up its own can of worms: short often means superficial, and superficial often means meaningless. There is safety in numbers - each category should have at least 2 questions assessing it, so if one of the questions is not working well, the other questions will compensate for it, and still provide accurate data. If you use short surveys, consider the following survey strategies to avoid the pitfall of superficiality:

Consider simplifying the construct. Short surveys do not measure complex constructs well, because there is too much ground to cover. For example, a long survey that covers multi-stakeholder issues, critical leadership behaviors, and job satisfaction can become a short survey by focusing in on one or two of those categories, instead of all three.

Focus on the most important categories. Short surveys can provide valuable, in-depth information if their questions focus on a few important issues. A thorough needs analysis featuring interviews and or focus groups can target critical areas needing special attention.

Use surveys in sequence. Short surveys featuring superficial coverage of lots of topics are appropriate provided there are follow-up surveys which "dive deep" in the problem areas which were identified.

Content Validity

Content validity tells you whether your survey construct is any good. Valid surveys are supposed to measure what they say they will measure. Content validity offers two tests of quality:

Should more content be included? Given what is known about the topic you are interested in from academia, industry and consulting, does this survey cover the bases? Are all the important aspects of this topic measured? Are there other aspects which need to be included for the survey findings to provide an accurate and comprehensive picture of this topic? If certain aspects are excluded, is there a reasonable and logical justification? The academic challenge has always been: "Why have you decided to include these aspects and not others?" If you do not have a good reason, you have just hung a target around your neck.

Example: A survey which claims to measure effective communication, and fails to assess whether dissenting opinions are invited and welcomed, is not covering the conceptual turf. If tolerating dissent is problem, this omission could be costly - managers thinking they are effective communicators when, in fact, people feel threatened and aren't talking. This is a content validity problem.

Should more content be excluded? Given what is known about the topic you are interested in from academia, industry and consulting, does this survey construct make sense? Are the categories focusing on the specific aspect they claim to measure, or are the questions in that category actually measuring any number of aspects? Finally, are any of the categories and questions not relevant to the topic, and should be tossed out altogether?

Example: A survey which uses job satisfaction questions to measure employee productivity has a content validity problem. Research has shown that happy employees are not necessarily productive employees. More job satisfaction questions need to be excluded and more questions on productivity topics such as efficiency must be added.

Assessing Content and Face Validity. Content validity, as well as Face Validity, which will be discussed next, are assessed by survey experts. In general, the more heads involved, the better. Each reviewer will offer different insights based on his or her experience, education, and expertise. PDi prefers the most rigorous standard in establishing content validity - the expert panel. An expert panel is typically composed of academics (preferably management professors), industry experts (consultants and/or industry analysts), and organizational practitioners (internal change agents from client firms). Expert panels maximize the diversity of perspectives brought to bear on refining and improving the survey construct.

Face Validity

Face validity answers two basic questions: Are respondents (1) able and (2) willing to give accurate answers to your survey? Each question is assessed to see whether its language, phrasing, and content is clear, unambiguous, unbiased, and easily understood. Face validity problems mean that respondents are likely to just guess at the answers, get mad and vent with their ratings, tell you what you want to hear, or transform the survey into a popularity contest. In any case, their ratings become impossible to interpret, and the data set useless, because you have no idea what was in their minds when they answered the question.

As the adage states: "Garbage in, garbage out."

Respondents are unable to answer when survey questions use phrasing they do not understand, or ask for knowledge they do not have. Typical pitfalls include:

Ambiguous phrasing, which could be interpreted any number of ways. For example, "Is your boss a good manager?" The term "good" is far too subjective.

Jargon, which is specialized terminology not generally understood. The term "personal engagement" may fall into this category, unless it is defined.

Verbose language, polysyllabic phraseology, inscrutably sesquipedalian in genesis.

Complex phrasing, where one question actually asks several questions. "My boss gives complete, actionable, detailed, timely feedback." Not 1 but 4 questions here.

Overestimation of understanding, where respondents are asked more than they know. "My boss often feels depressed." This is not assessment, it is psychoanalysis.

Respondents are often unwilling to answer accurately when the question has an obviously "correct" or "incorrect" answer. Most people want to associate themselves and their friends with positive, socially desirable attitudes and behaviors (the "halo" effect), and associate their competitors and enemies with negative, undesirable attitudes and behaviors (the "horn" effect) -- regardless of whether or not this characterization is true. This unfortunate rating propensity is called "demand bias," and is a major contributor to measurement error. Self-monitoring questions are notorious in this regard, since most respondents will rate themselves above average or higher on any given behavior -- a tendency known as inflated self-efficacy. Typical pitfalls include:

Inflammatory language, which makes some answers obviously wrong. Avoid creating a negative emotional response, unless you truly are stupid and incompetent.

Leading phrasing, which makes some answers obviously right. I am sure that managers of your intelligence and expertise would never do such a thing, right?

Loaded questions, which give hints and reasons favoring a certain answer. In light of your upcoming performance review, you do agree with your boss that state-of-the art, leading edge survey research creates vital transformations, don't you?

Pressures for demand bias become particularly intense if respondents feel that honesty might hurt them. Data will not be accurate if respondents fear that survey results may come back and bite them on their behinds. Unfortunately, two problems are common:

Buying Responses. When survey respondents know that the results will be used for compensation purposes (determining bonuses, promotions, or other perks), respondents are much more likely to give responses which favor themselves and their friends, and undermine their enemies - regardless of whether such ratings have much to do with the reality the questions hope to assess.

Killing the Messenger. When survey respondents fear efforts will be made to identify the "troublemakers" and "non-team players" who dared to give unfavorable responses, they will tend to use their surveys to tell management what management wants to hear. Why put your job at risk like "John Doe" did. He was honest, and came in to his next performance appraisal to see his boss thumbing through his survey, which had his name written on it.

Fortunately, there are some proven survey research strategies which tend to minimize demand bias:

Use survey information for development purposes. Respondents will be more accurate if they know this information is for positive, proactive change efforts, and will never be used for blaming, scapegoating or punishing.

Guarantee complete confidentiality, with the promise that only general trends shared by many

employees will be reported. For research purposes, unique individual responses are neither relevant nor interesting to the study.. Also guarantee complete confidentiality of the written responses, meaning that all names and specifics which could identify an individual employee or manager will be deleted or otherwise appropriately edited.

Limit the number of demographic questions to 5 or under. More demographics raise the suspicion that respondents can be tracked down and identified. (How many Asian, female, marketing managers at location 2 with 5 years of tenure with a Ph.D. are there?)

Only report on demographic comparisons containing more than 10 employees.

If trust is low, **have the survey distributed and collected by an independent firm**, which will keep the survey forms and only distribute anonymous data sets.

Construct Validity

A survey can be validated in several ways, and on several different levels. This outline of the survey process will conclude by reviewing all of the components of construct validity - briefly summarizing those already covered, and explaining the more advanced techniques. Individual questions can be tested for face validity, and for a normal distribution around the mean. Categories and sub-categories can be tested for content validity and for reliability. The survey can be tested for construct validity, involving discriminant, convergent and predictive (criterion) validity.

Individual Questions

After testing for face validity, the questions can administered and tested for a normal distribution of data around the mean. These tests identify three common problems, all of which can be statistically screened by computing skewness and kurtosis statistics:

Skewness, where most respondents are giving extremely high or extremely low ratings.

Restricted Variance, where most respondents are giving the same rating.

Flat or Bimodal Distributions, where respondents are giving more high and low ratings than expected.

Categories and sub-categories

Categories or sub-categories can be tested for content validity before the administration, and tested for reliability afterwards by computing a Cronbach Alpha reliability coefficient.

Survey Construct

Once the sub-categories and categories have been validated and tested for reliability, they become variables which can be used to test the survey construct, using statistical techniques such as Pearson correlation and inter-correlation matrices, and factor analysis. These techniques test the relationships between variables to see if they are independent, and if they have a significantly positive or negative relationship.

Convergent Validity. Some categories and sub-categories are expected to correlate positively or negatively with each other. For example, trust, openness, and cooperation represent categories which should share a positive relationship - high ratings in any one category should be matched by high ratings in the other. In contrast, backstabbing represents a category which should share a negative relationship with openness -?high ratings in one category should be matched by low ratings in the other. Convergent validity forces researchers and change agents to challenge their assumptions when expected positive correlations are not there, or negative correlations pop up where they were not supposed to be. The results of the hypothesis testing should also be consistent with the construct. In its most rigorous sense, Convergent validity means that not only do the survey categories and sub-categories relate to each other as expected, but also that information from a variety of sources besides surveys also triangulate and confirm the survey findings as expected. Thus data from different sources support each other, and "triangulate" on a research finding.

Discriminant Validity. While some categories are expected to yield similar ratings, they should not be almost identical. When they are too similar, they are not providing unique information, and become

redundant and superfluous. In short, we need each category to tell us a different story, or it is a waste of space. Exploratory Factor Analysis is often used to establish discriminant validity, because it collapses questions into the fewest number of distinct factors as possible. Redundant categories will load together on the same factor, and can be safely condensed. Unanticipated factor loadings also can identify unexpected categories, which can be used if they make conceptual sense.

Predictive (Criterion) Validity. In some cases, surveys are used to predict actual performance. In this case, concurrently with the survey administration, performance data is gathered on the individuals or groups being assessed. With these data, the final step in the survey analysis is the use of statistical tests to determine whether the survey variables (developed from the categories and questions) are significantly associated with the performance data.

The Validation Process

The goal of construct validity is simple: to design survey instruments that deliver as promised. The more extensive and rigorous the validation process, the more confidence you can have in the power and relevance of the survey results. Ideally, the validation process is extremely rigorous, and follows the following sequence:

Comprehensive Needs Analysis - including extensive interviews and focus groups.

Face and Content Validity - assessed by a large expert panel from a variety of professional backgrounds, including academics, industry experts and practitioners.

Face Validation Pretesting - a survey about the survey, specifically asking if each question is understandable and matches the construct it is supposed to measure.

Face Validity Focus Groups - after taking the Face Validity survey, further comments are solicited and discussed on each question, as well as suggestions for new questions.

Validity and Reliability Pretesting - after Face Validation, but before the main administration the survey is given to a smaller, representative sample of the organization, and statistically analyzed. Revisions are made accordingly.

Validity and Reliability Posttesting - after the main administration, the data set is analyzed for Construct Validity, Reliability, and Predictive (Criterion) Validity (if desired). Hypothesis testing typically adds to the power of the research findings.

Convergent Validity - Research findings are systematically compared with other types of data and information which were gathered concurrently with the survey effort.

Iterative Survey Refinement - The feedback from the statistical analysis and management debriefs is used to improve individual questions, to change categories and demographic variables, and to shorten the instrument (if possible). The effectiveness of such changes can be explored during subsequent iterations.

Many organizations lack the time, resources and energy for a rigorous validation process. Here are the strengths and weaknesses of common validation strategies:

Face and Content Validity Without Inferential Statistics. This strategy is clearly the most cost effective, since it does not involve statistical analysis and pre-testing. Further, provided the expert panel has genuine expertise, they can often effectively manage or avoid typical problems and pitfalls in survey design and construction which they have resolved before. The effectiveness of this approach rests on the quality of the expert panel, as well as the thoroughness of the needs analysis data (interviews, focus groups, etc.) that led to the survey.

However, no matter how expert your panel, they are not a psychic hotline, and can not avoid all problems with question phrasing and survey design. There are always surprises, and those problems cut into the effectiveness and validity of the survey findings. Many questions may have to be reworded, discarded, or added to subsequent iterations as problems in their phrasing, or as problems

in the construct become apparent. Further, without calculations of statistical significance, whatever differences, trends, and patterns emerge from the data can not authoritatively be distinguished from random error, coincidence, or chance. Finally, descriptive statistics alone represent only a fraction of the explanatory power of the data, which will be lost without inferential statistical analysis.

Validate the Survey by Pre-testing. This strategy goes beyond face and content validity, and takes the survey out for a test run to work out any bugs in the design. The results of the pre-test become the basis for another revision and results in the final version of the survey. The advantage of this approach is that there are few surprises and unanticipated problems in the actual administration, maximizing the effectiveness and validity of the subsequent research findings. This approach is particularly useful if one of the goals is to establish benchmarks for future comparison C pretested questions have a much better chance of holding up under subsequent statistical analysis, and are less likely to require the kind of extensive rewording which would make them invalid for benchmark comparisons. The effectiveness of this approach rests on the quality of the pre-test.

The downside is additional time and money. The larger and more representative the pre-test, the more expensive and time consuming it becomes. Smaller pre-tests using respondents who are easily accessible and convenient are less expensive and time consuming, but "samples of convenience" are notorious for not being representative of the organization, or particularly accurate in their findings. Also, pre-testing is not an adequate substitute for rigorous face and content validity testing C poor surveys may require several pre-tests before they are sufficiently refined.

If pre-testing is not followed by statistical analysis after the administration, you are not taking advantage of your validated data. Without calculations of statistical significance, whatever differences, trends, and patterns emerge from the data can not authoritatively be distinguished from random error, coincidence, or chance. Finally, descriptive statistics alone represent only a fraction of the explanatory power of the data, which will be lost without inferential statistical analysis.

Validate the Survey by Post-testing. This strategy uses face and content validity testing before the administration, and uses the first administration as a large-scale pretest. Periodic administrations become iterative design tools, with the survey evolving as each set of results identifies further areas for refinement. The advantage of this approach lies in avoiding the costs and time of the pre-test -- validation can be combined with hypothesis testing after the administration. The effectiveness of this approach rests on the quality of the expert panel, as well as the thoroughness of the needs analysis data (interviews, focus groups, etc.) that led to the survey.

The downside is there will be problems with the first few iterations of the survey, and those problems cut into the effectiveness and usefulness of the survey findings. Whatever time is saved by avoiding a pre-test will be lost by delays waiting for the results of subsequent administrations C many questions may have to be reworded, discarded, or added to subsequent iterations as problems become apparent. These issue have to be resolved before a company benchmark can be established.

Implications

Surveys have amazing potential. Carefully crafted survey instruments can accurately measure and report on the attitudes and behaviors of large groups of employees more cheaply and effectively than virtually any other method of data collection. The challenge lies in the construction, for designing effective survey instruments is very tricky. In fact, it is so challenging that despite the best efforts of experienced internal and external change agents, surveys often do not perform as expected. When surveys are not all they are supposed to be, at best some of the data is useless, representing a waste of time, effort and money. At worst, some of the data is inaccurate and potentially misleading those who trust in the survey findings. And the scariest thing about surveys is that there is no such thing as the "perfect survey." Survey construction is as much of a balancing act as a science. Consider the following research questions:

Given the topics of interest, is this survey too short or too long?

Is the survey too in-depth on too few topics, or too superficial on too many?

Are questions yielding actionable data, or are they unanalyzed abstractions, wandering generalities or meaningless specifics?

Given the purpose and intended use of the survey, what level of validation is necessary?

Can we hold people accountable for survey results without statistical tests of significance? On which topics, and why?

Can I afford to pretest the survey? Can I afford not to?

Clearly this process of survey research is time consuming, complex and difficult to do well. For those change agents who would like to improve the efficiency and effectiveness of their survey process, Performance Dimensions International, LLC, offers a wide variety of services and resources.

Contact PDi about these publications

(copyright 2002 by PDi. All rights reserved. Please do not modify, copy or distribute this article -- see website [terms and conditions of use.](#)) 