

**WHITE PAPER ON NORMS**

**PREPARED FOR:**

**DENISON CONSULTING**

**MARCH 2007**

**Prepared by:**

**Mike Zickar, PhD  
Patricia Barger, MA  
Ashley Guidroz, MA  
Maya Yankelevich, BA**



**Institute for Psychological Research and Application  
DEPARTMENT OF PSYCHOLOGY  
BOWLING GREEN STATE UNIVERSITY**

Survey results, presented by themselves, are difficult to interpret. There is a growing body of literature on how to make such scores more easily interpretable, especially for consumers of survey tools who are untrained in psychometrics and statistics. Although there are many different techniques for making survey results more intuitively meaningful, there are many contradictions in the test interpretation literature. In addition, although there has been much written about test interpretation and norms, much of it is scattered among various disciplines with little systematic treatment. There has been very little comprehensive treatment about the different statistics used in creating norms, and their advantages and disadvantages

### *Background Information*

When interpreting the results of organizational surveys, managers may be interested in knowing their organization's relative, as opposed to absolute, standing on a topic of interest (e.g. employee satisfaction). Interpreting survey results in terms of relative standing provides information about where an organization stands *in relation* to similar organizations. On the other hand, interpreting survey results with regard to absolute standing simply tells an organization how they are performing within *their own* organization; thus, there is no broader context or frame-of-reference to evaluate the absolute standing results. Thus, the goal of an organizational survey should be to determine an organization's relative standing on a topic of interest so that comparisons can be made to other organizations within or across industries.

In this white paper, we review different techniques to increase the interpretability of survey results. After reviewing key definitions and techniques, we describe advantages and disadvantages of each of the techniques. Next, we analyze a large data set to evaluate some of the key metrics. Finally, at the end, based on our review of the literature, we provide specific

recommendations for increasing the interpretability of survey results for consumers while maintaining the scientific integrity of those test norms.

### *Key Definitions and Techniques*

We present three common metrics by which organizations judge how well they are doing on constructs of interest and explain why or why not these metrics are appropriate. The three metrics we discuss are mean scores, percent favorable responses, and percentiles (benchmarks). First, we will define each of these and then provide evidence for their applicability.

#### *Definitions*

*Mean scores.* A mean is the most basic of descriptive statistics. It is an average response for a given item obtained across all respondents. Mean scores are calculated by summing each individual's response (for example: 5, 3, 5, 3, 4 = 20) and then dividing by the number of scores (e.g., 5, or the number of people responding). In this case the mean score is 4 (20 divided by 5). Mean scores are simple yet effective descriptive measures and have a wide range of application and usefulness (Rogelberg, Church, Waclawski & Stanton, 2002). Although there are problems with doing this, mean scores are often used to identify the highest and lowest-rated items (Church & Waclawski, 1998). It is also easy to sort questions by rank ordering their mean scores from highest to lowest. As we discuss later, however, means have many limitations that can hinder the interpretability of test scores.

*Percent favorable.* Percent favorable is the frequency of individuals who respond favorably (positive response) to any given question. For example, let us examine what the percent favorable score would be for a group of 10 people responding to a survey item on a 5-point scale, with '1' being the most negative and '5' being the most positive. Out of ten respondents one individual responded with a '4', and four individuals responded with a '5', then

the percent favorable for that particular question would be 50% (five out of 10 responding favorably). Percent favorable presents frequency of responses in a collapsed and simplified version. Later in the paper we will discuss limitations of this approach.

*Benchmarking/Percentiles.* Percentiles are percentages that are created through the use of a normative sample. A percentile indicates the percentage of individuals in that sample who are at or below the given score. For example, if a department scores an average of 3.92 on a measure of Core Values, the test norms developed for this scale can be used to determine what percentile the department falls in. Referencing the norms may indicate that a score of 3.92 translates to the 92<sup>nd</sup> percentile in Core Values. In other words, the department's perception of the organization's core values is higher than 92% of other departments surveyed. Additionally, the department's score can also be compared to other departments within their organization or other organizations within their industry. Thus, percentiles are dependent on their normative group, or the group that is used for comparison. In this lies the advantage of percentiles: they provide users with a comparison group. Percentiles are advantageous because they provide users with the ability to make both internal and external comparisons. As will be demonstrated, however, the utility of norms is inherently tied to the relevance of the comparison group.

#### *Evaluations of the Different Techniques*

##### *Mean Scores*

*General Characteristics of Means.* Mean scores can be advantageous for comparing absolute scores across workgroups on different items of interest within the company, provided the items are worded as similarly as possible and each item has a similar response scale, such as a 5-point scale (Church & Waclawski, 1998). Because the wording of the items can drastically influence the mean, as well as other statistics, using items with consistent stems is important (i.e.

all items begin with the phrase “I believe that my organization....”). Assuming that items are similarly worded and on a similar response scale, an organization can use mean statistics to make comparisons on single items across departments as well as make comparisons between two items.

*Comparing Means Across Departments.* Comparing means across workgroups may provide the organization with information as to which departments are in most need of improvement in particular areas of interest. For example, a company can compare a mean rating of 4.5 on a ‘supervision’ item in the marketing department to a 3.8 mean rating on a ‘supervision’ item in the accounting department, and conclude that, within that particular organization, accounting employees are relatively more satisfied with their supervision than are the marketing employees. It is important to note that the types of inferences that can be made from comparing means across workgroups must be interpreted with caution, even given that the items have similar wording. Even if a particular department obtains the highest mean score on a particular item, it does not necessarily indicate that that department does not need improvement in that area. The highest scoring department may in fact actually have a *low* score overall when compared to *other organizations*, and so it is important for organizations to keep in mind that mean information is only interpreted relative to their individual organization and does not provide information about a department’s standing compared with outside companies.

*Disadvantages to Means.* Although means can provide useful information about the absolute standing of various items across departments within the organization, there are some potential problems in using means to interpret survey results. Specifically, disadvantages to using means lie in their susceptibility to statistical outliers or skewed data, the limited inferences that

can be made when comparing between items, and the inability to make comparisons to other organizations by themselves.

Using a mean statistic to interpret survey results invites the opportunity for misinterpretation, especially when outliers are present or when the sample is not normally distributed. Outliers are responses that are extreme when compared with the rest of the data (i.e. 1, 1, 2, 3, 278). Outliers can distort means and cause them to lose their representativeness. Furthermore, the interpretation of means can be deceptive when the data has a bimodal distribution (i.e. containing two distinct values that the data tend to center around), or when the data is skewed (i.e. when the data is not symmetrically distributed around the mean) (Rogelberg, Church, Waclawski & Stanton, 2002).

An additional disadvantage to using means is that comparisons made between items provide little insight into the responses beyond comparing across workgroups. Church and Waclawski (1998) argue that means can be used to rank-order items in order to identify strengths and weaknesses. However, rank-ordering is not recommended because this comparison assumes that all items have the same amount of item difficulty. Item difficulty, in this sense, refers to how an item is worded and is influenced by the use of adjectives. Take, for example, the following items: “My supervisor is an effective leader” or “My supervisor is a *highly* effective leader.” The latter item is more difficult because endorsement of that item implies a higher level of performance of the supervisor. In reading the first question, an employee with an average supervisor and an employee with an above average supervisor would probably respond the same way. However, if asked if their supervisor is highly effective the employee with the average supervisor probably would not endorse this item as strongly. Hence, rank-ordering items as strengths and weaknesses can be problematic for interpretation, particularly if some of the items

are worded in a way that makes it easier to get a favorable response. Thus, as mentioned before, when using means one needs to pay particular attention to the way the item was written and interpreted by respondents.

Perhaps one of the most important disadvantages of mean scores is that there is no standard for assessing how the organization is doing *compared to similar organizations* (Edwards, Thomas, Rosenfeld & Booth-Kewley, 1997). Although comparing across groups for items *within a single organization* is possible with means, determining an organization's standing on a single item is difficult. For example, if an organizational survey revealed a mean score of 4 on a 5-point scale for employee satisfaction, what does that mean? Although it may lead managers to believe that their employees are above-average in their feelings of satisfaction, this is a deceptive conclusion. What if the average level of satisfaction for all employees in the U.S. in the organization's industry was 4.5? In light of this information, clearly, the organization's mean score of 4 now does not appear to be 'above average' – in fact, it is *below average*. Thus, when attempting to determine the relative standing of an organization, mean scores are not a useful method to interpret survey results.

#### *Percent Favorable*

Percent favorable is the most common approach for reporting survey results (Rogelberg, Church, Waclawski & Stanton, 2002). One advantage of using percent favorable to report results is that it collapses responses into simple groupings. For example, on a 5-point culture scale, '4' (i.e., agree) and '5' (i.e., strongly agree) would be collapsed into one category indicating a "favorable" response. In this way, percent favorable simplifies the data and presumably make it easier to communicate the data results (Jones & Bearley, 1995). Despite its simplicity, the percent favorable approach to data analysis has many disadvantages that caution

against its use. First, by collapsing continuous data into dichotomous data (i.e., two categories of ‘favorable’ vs. ‘unfavorable’), the percent favorable approach ignores the advice of standard measurement theory to use continuous data whenever possible (Nunnally, 1978). When dichotomizing continuous data, much valuable psychometric information is lost (Edwards, Thomas, Rosenfeld & Booth-Kewley, 1997), including reduced variability and reduced discrimination that respondents made among the categories when initially completing the survey.

Although collapsing data makes interpretation easier, the results can be misleading. For example, if we take a set of 20 responses to a culture item where 15 people responded favorably and five people responded unfavorably, according to the percent favorable approach, the general consensus would be that the department has a strong culture. However, taking a closer look at the data we might find that out of the 15 favorable responses, only two are a score of ‘5’ (i.e., strongly agree) and the rest are ‘4s’ (i.e., agree). Similarly, out of the five unfavorable responses, all of them could be scores of ‘1’ (i.e., strongly disagree). Thus, the results are in fact much more negative than initially perceived. The level of intensity is neglected when reporting percent favorable.

In addition to the misleading qualities of such reporting mechanisms, collapsing data puts a new set of restrictions upon the underlying structure of the dichotomous scale that did not exist in the continuous scale (Rogelberg, Church, Waclawski & Stanton, 2002). These problematic psychometric properties can seriously compromise the quality of interpretations of the data. Thus, although the intent of percent favorable scores is to simplify and clarify survey results, the consequence of such a technique can result in misinterpretation (Rogelberg, Church, Waclawski & Stanton, 2002). Additionally, when using percent favorable responses the tendency is to then rank all the items from most to least favorable and only focus on improving the least favorable



items, thus ignoring valid information, particularly when items have different average scores across the population of interest (Macey & Eldridge, 2006).

### *Benchmarking/Percentiles*

The percentile method of data reporting circumvents some of the weaknesses associated with the mean score and percent favorable approaches because of the capacity to provide both internal and external comparisons. Recall that using percentiles is a form of ‘benchmarking,’ whereby organizational data are compared to a normative sample. This normative sample can be created through the use of external norms, such as a national sample of organizations, or internal norms, such as norms based on comparison between departments within one organization or benchmarked within one organization across years. The advantage of benchmarking is that it allows interpretation of data by comparing survey results to the performance of companies outside of the organization or across departments within the same organization (Rogelberg, Church, Wacalawski & Stanton, 2002).

A benchmark uses a percentile which indicates the percentages of groups or organizations that are below an organization’s score. Thus, a survey that revealed an aspect of an organization’s culture as being in the 75<sup>th</sup> percentile would indicate that 75% of similar organizations are equal or lower than the target organization on that culture attribute. Although mean scores and percent favorable data do not have much inherent meaning because they lack a frame-of-reference within which to be interpreted, percentiles allow organizations to see their results in context (Rogelberg, Church, Wacalawski & Stanton, 2002). Thus, when using percentiles, an organization is able to obtain their relative standing compared to other organizations on a question of interest, and thus will know how they compare to that normative sample (Macey & Eldridge, 2006). An advantage of this information is that it can create energy

for action, whereby managers strive to be among the top performers in their industry. Knowing where an organization stands relative to others also helps managers focus on the areas where they are weak and thus helps them avoid spending scarce resources in areas where they are already doing well compared to others (Macey & Eldridge, 2006). Without benchmarking information, the only assumption a manager has to work with is that a higher mean score or percent favorable is better.

However, benchmarking is not without its criticisms. Arguments against the use of benchmarking center around the ease of use of benchmarks and the choice of a normative sample. Percentiles can be misinterpreted because untrained test users have a natural tendency to interpret percentiles as absolute scores. For example, “if a person scores 5 percentile points above another person, the latter must be a 5 percent better person for this job!” (Lawshe & Balma, 1966, pg. 75). This is a tenuous conclusion because most tests are not reliable or valid enough for a difference of 5 percentile points to translate to an actual difference in ability (i.e. it may have resulted from pure chance or unreliability in the survey administration). The concept of percentiles is complex, and not subject to simple adding and multiplicative properties. As previously mentioned, the distribution of the percentile mirrors that of the raw sample so a percentile score of 60 may not be twice as good as a percentile score of 30 (Lawshe & Balma, 1966).

A second argument against the use of benchmarks is in regard to the choice of the normative sample. A conservative view of organizational benchmarking argues that no external comparisons should be made, rather organizations should instead benchmark to the ‘ideal’ (Rogelberg, Church, Waclawski & Stanton, 2002). Proponents of this perspective contend that companies should only be concerned with where they want to be and not with where they are in

comparison to other organizations. However, if an organization chooses a normative sample to benchmark on, these comparisons can be made internally, within the company, or externally, to a large sample of other organizations. One criticism of making external comparisons is that the normative sample may not relate to the organization in terms of organizational characteristics such as type of company, industry, or demographic make-up of employees (Lees-Haley and Lees-Haley, 1982). From an equivalence perspective, benchmarks are not as valid when the groups used to compare are not suitably equivalent. Thus, when interpreting benchmarking data, it is also important to know the characteristics of the normative sample to which the comparisons are being made (Rogelberg, Church, Waclawski & Stanton, 2002). What is most important in the use of benchmarks is for an organization to evaluate the possible benchmark comparisons that are available to them and choose the comparison that will provide the most valuable information to them for meeting their organizational goals.

#### *An Empirical Evaluation*

The following section will illustrate the advantages and disadvantages of using each statistic when interpreting survey results using data from the Denison Organizational Culture Survey (DOCS). As discussed previously, percentiles offer a number of advantages over using means or percent favorables when interpreting organizational survey data. The mean is a useful statistic because it provides an organization's standing on different items of interest and is also a robust measure of central tendency. As an example, Table 1 includes the average score of data collected on 7 items from the DOCS. This organization scored highest in employee involvement (Item 1) and core values (Item 2) and lowest on innovation (Item 6) and coordinating projects across levels of the organization (Item 7). Without a context for interpreting these scores, an organization could conclude that more effort should be directed at rewarding innovation among

their employees and improving procedures for coordinating and integrating projects across departments. This decision could mean that little attention would be directed toward improving the level of involvement of employees or strengthening the core values of the organization. This decision, however, would be misinformed because all the scores in Table 1 are at the 50<sup>th</sup> percentile of the DOCS norms. That is, taking into consideration that although each score is numerically different, the mean scores for each item are no better than 50% of the organizations that have all completed the DOCS. So, even though the mean score for involvement (Item 1) was much higher than the mean score for coordinating projects (Item 7), relatively speaking, they are the same when compared to other organizations. Thus, although the mean captures a company's average response to an item, it provides little information beyond that. There is no context for interpreting whether the mean score is "good" or "bad" without the addition of percentiles.

Table 1. Average Score of Example Organization on 7 DOCS Items

<b>Survey Item</b>	<b>Mean Score</b>
1 Most employees are highly involved in their work.	3.97
2 Ignoring core values will get you in trouble.	3.80
3 We continuously track our progress against our stated goals.	3.63
4 New and improved ways to do work are continually adopted.	3.39
5 The leaders and managers "practice what they preach."	3.20
6 Innovation and risk taking are encouraged and rewarded.	3.04
7 It is easy to coordinate projects across different parts of the organization.	2.75

Similarly, percent favorables can also provide useful information to an organization, but this method has some limitations as well. On the DOCS, the percent favorable is calculated by collapsing the number of employees who responded with a 4 or 5 on an item. Calculating a percent favorable is valuable in that it indicates what percentage of the company endorsed an item favorably. Continuing with the previous example, our organization scored at the 50<sup>th</sup> percentile on 7 DOCS items, but the overall averages for each item correspond to different

amounts of percent favorable and mean scores (See Table 2). The percent favorable for item 1 was 86%; however, the percent favorable for item 7 was 40%. Relying only on percent favorable or the average score on an item can be misleading and can hamper organizations in their development efforts.

Table 2. Average Score, Percentile, and % Favorable of an Example Organization on 7 DOCS Items

	<b>Survey Item</b>	<b>Mean Score</b>	<b>Percentile</b>	<b>% Favorable</b>
1	Most employees are highly involved in their work.	3.97	50	86%
2	Ignoring core values will get you in trouble.	3.80	50	82%
3	We continuously track our progress against our stated goals.	3.63	50	86%
4	New and improved ways to do work are continually adopted.	3.39	50	64%
5	The leaders and managers “practice what they preach.”	3.20	50	57%
6	Innovation and risk taking are encouraged and rewarded.	3.04	50	41%
7	It is easy to coordinate projects across different parts of the organization.	2.75	50	40%

Percentiles are far more advantageous in survey research because they provide a context for interpreting results. Percent favorable captures what proportion of respondents agree or strongly agree with an item and a mean score captures the average overall. Both of these statistics provide some useful information and are good indicators of overall agreement with an item. However, a standard is still needed to help interpret the meaning of these scores. Table 2 illustrates that large differences in mean scores and percent favorable responses imply little without the context to interpret the scores. The DOCS norms were created using percentiles which provide a context to interpret what an organization’s score means relative to other organizations in the database.

### *Specific Recommendations*

In light of the preceding discussion of the strengths and weaknesses of the various metrics used to interpret survey results, it is recommended that organizations use benchmarking (percentiles) whenever possible. Overall, percentiles offer improved interpretation of survey data because they provide the organization with information about its *relative* standing on items of interest. This allows for useful comparisons between organizations and/or departments within a single organization. Mean scores and percent favorables do not lend themselves to this type of interpretation and are not recommended when the percentile method is available.

To glean the most useful information from a percentile metric, it is important to use survey items that have existing normative data. Large databases of data are available that allow for the creation of benchmarks. As such, it is recommended that as few of changes as possible be made to a survey so that the normative databases may be of use in the survey interpretation process. Adding unique items to existing scales or deleting unwanted items from the scales may alter the psychometric properties of the original scale. This will make comparisons to benchmark data difficult because benchmarking assumes identical items are given across organizations. Altering existing scales may hinder the interpretation of survey results and contradict the primary advantage of using a percentile method.

If an organization elects to include unique items in a survey, the resulting statistics should be interpreted cautiously. As mentioned, the organization will be unable to compare this data with normative data, but may look to mean scores for information, if done with caution. When using mean scores to interpret items unique to a particular organization, it is recommended that the organization examine the distribution of scores to ensure a normal, uni-modal distribution with few, if any, outliers in the data set. It is also important to make sure that during item

development close attention is paid to the wording of items, as wording can have an effect on responses. Specifically, the unique items should adopt similar item stems as present in the existing measure and should also use the same response options. When these conditions are satisfied the interpretation of mean scores would be acceptable, as this will minimize the risk of misleading results.

The use of percent favorable is not recommended under any conditions. Utilizing the percent favorable method causes the loss of important psychometric information leading to possible misinterpretation of survey results. Despite its simplicity, percent favorable should not be used. In conclusion, it is highly recommended that organization utilize survey items with existing normative data. Additionally, when the situation requires the addition of unique survey items it is recommended that mean scores are used for interpretation.

Table 3. Limitations and Strengths of Mean Scores, Percent Favorable, and Percentiles

<b>Limitations</b>	<b>Strengths</b>
<b>Mean Scores</b>	
<ul style="list-style-type: none"> <li>▪ No context. Thus, interpretation of rank-ordered scores must be done with caution</li> <li>▪ Comparing items using mean scores provides no insight into responses beyond their rank</li> <li>▪ Overly influenced by outliers</li>   <li>▪ Bimodal and skewed distribution of means makes interpretation ambiguous</li> <li>▪ No standard for assessing how organization is doing compared to similar organizations</li> </ul>	<ul style="list-style-type: none"> <li>▪ Robust measure of central tendency</li> <li>▪ Simple, wide range of application</li>   <li>▪ If norming data are not available: good for comparison and rank ordering of absolute scores on similarly worded items within an organization</li> </ul>
<b>Percent Favorable</b>	
<ul style="list-style-type: none"> <li>▪ Does not comply with standard measurement theory which advises to use continuous data</li> <li>▪ Valuable psychometric information is lost</li> <li>▪ Reduced variability of scores</li> <li>▪ Reduced discrimination of responses made by respondents</li> <li>▪ Results are misleading</li> <li>▪ New set of statistical restrictions placed on the collapsed scale, compromising accurate interpretation of results</li> <li>▪ Valid psychometric information ignored when percent favorables are ranked</li> </ul>	<ul style="list-style-type: none"> <li>▪ Collapsed and simple presentation</li> </ul>
<b>Percentiles</b>	
<ul style="list-style-type: none"> <li>▪ Not as accurate when comparison groups are not equivalent. Must take normative group into consideration. Norm data must be relevant to the comparison being made</li> <li>▪ Compares organizations with other observed data, not what is ideal</li> <li>▪ Untrained test users interpret percentiles as absolute scores</li>   <li>▪ Not subject to simple adding and multiplicative properties</li> </ul>	<ul style="list-style-type: none"> <li>▪ Addresses weakness of mean scores and percent favorable scores</li>   <li>▪ Provides comparison group—internal (within the organization) or external (other organizations)</li> <li>▪ Provides insight into whether score differences are normal or not. That is, is the organization's score normal compared to other organizations</li> <li>▪ Provides energy for action</li> </ul>



## References

- Church, A.H & Waclawski, J. (1998). *Designing and using organizational surveys: A seven step process*. San Francisco, CA: Jossey-Bass.
- Edwards, J.E., Thomas, M.D., Rosenfeld, P. & Booth-Kewley, S. (1997). *How to conduct organizational surveys: A step-by-step guide*. Thousand Oaks, CA: Sage.
- Jones, J.E. & Bearley, W.K. (1995). *Surveying employees: A practical guidebook*. Amherst, MA: HRD Press.
- Lawshe, C.H. & Balma, C.J. (1966). *Principles of personnel testing* (2<sup>nd</sup> ed). New York, NY: McGraw Hill.
- Lees-Haley, P.R. & Lees-Haley, C.E. (1982). Attitude survey norms: A dangerous ally. *Personnel Administrator*, 89, 51-53.
- Macey, W.H. & Eldridge, L.D. (2006). National norms versus consortium data: What do they tell us? In A.I. Kraut (Ed.) *Getting action from organizational surveys: New concepts, Technologies, and applications* (pp. 352-376). San Francisco: Jossey-Bass.
- Nunnally, J.C. (1978). *Psychometric theory* (2<sup>nd</sup> ed.). New York: McGraw Hill
- Rogelberg, S.G., Church, A.H., Waclawski, J. & Stanton, J.M. (2002). Organizational survey research. In S.G. Rogelberg (Ed.) *Handbook of Research Methods in Industrial Organizational Psychology*. Malden, MA: Blackwell Publishing Ltd.