

# THE GALLUP ORGANIZATION

## **Q<sup>12</sup><sup>®</sup> Meta-Analysis**

**James K. Harter, Ph.D.**

The Gallup Organization

**Frank L. Schmidt, Ph.D.**

University of Iowa

**Emily A. Killham, M.A.**

The Gallup Organization

**James W. Asplund, M.A.**

The Gallup Organization

March 2006

THE GALLUP ORGANIZATION

1001 Gallup Drive  
Omaha, Nebraska 68102

# Copyright Standards

This document contains proprietary research, copyrighted materials, and literary property of The Gallup Organization. It is for the guidance of your company only and is not to be copied, quoted, published, or divulged to others outside of your organization. Gallup<sup>®</sup>, SRI<sup>®</sup>, and Q<sup>12</sup><sup>®</sup> are trademarks of The Gallup Organization, Princeton, NJ. All other trademarks are the property of their respective owners.

This document is of great value to both your organization and The Gallup Organization. Accordingly, international and domestic laws and penalties guaranteeing patent, copyright, trademark, and trade secret protection protect the ideas, concepts, and recommendations related within this document.

No changes may be made to this document without the express written permission of The Gallup Organization.

## Acknowledgments

---

The authors thank Mark Stiemann for his editing expertise. We also thank numerous Gallup scientists who have completed studies incorporated into this ongoing meta-analysis.

# Contents

<b>Introduction</b>	<b>4</b>
<hr/>	
Foreword .....	4
Development of the Q <sup>12</sup> .....	5
Introduction to the Study .....	7
Description of the Q <sup>12</sup> .....	9
<b>Meta-Analysis, Hypothesis, Methods, and Results</b>	<b>14</b>
<hr/>	
Meta-Analysis .....	14
Hypothesis and Study Characteristics .....	14
Meta-Analytic Methods Used .....	21
Results.....	24
<b>Utility Analysis: Practicality of the Effects</b>	<b>30</b>
<hr/>	
<b>Discussion</b>	<b>36</b>
<hr/>	
<b>References</b>	<b>38</b>
<hr/>	
<b>Appendices</b>	<b>43</b>
<hr/>	
<i>Appendix A: Reliabilities of Business-Unit Outcomes .....</i>	<i>43</i>
<i>Appendix B: Test-Retest Reliabilities of Employee Engagement.....</i>	<i>44</i>

# Q<sup>12</sup>® Meta-Analysis

## *Introduction*

### **Foreword**

---

In the 1930s, George Gallup began a worldwide study of human needs and satisfactions. He pioneered the development of scientific sampling processes to measure popular opinion. In addition to his polling work, Dr. Gallup completed landmark research on well-being, studying the factors common among people who lived to be 95 or older (Gallup & Hill, 1959). Over the next decades, Dr. Gallup and his colleagues conducted numerous polls throughout the world, covering many aspects of people's lives. His early world polls covered topics such as family, religion, politics, personal happiness, economics, health, education, safety, and attitudes toward work. In the 1970s, Dr. Gallup reported that less than half of those employed in North America were highly satisfied with their work (Gallup, 1976). Work satisfaction was even lower in Western Europe, Latin America, Africa, and the Far East.

Satisfaction at work has become a widespread focus of researchers. In addition to Dr. Gallup's early work, the topic of job satisfaction has been studied and written about in more than 10,000 articles and publications. Because most people spend a high percentage of their waking hours at work, studies of the workplace are of great interest for psychologists, sociologists, economists, anthropologists, and physiologists. The process of managing and improving the workplace is of great importance and presents great challenges to nearly every organization. So, it is vital that the instruments used to create change do, in fact, measure workplace dynamics that predict key outcomes — outcomes that a variety of organizational leaders would consider important. After all, organizational leaders are in the best position to create interest and momentum around job satisfaction research.

Parallel to Dr. Gallup's early polling work, Donald O. Clifton, a psychologist and professor at the University of Nebraska, began studying the causes of success in education and business. Dr. Clifton founded a company called Selection Research, Incorporated in 1969. While most psychologists were busy studying dysfunction and the cure of disease, Dr. Clifton and his colleagues focused their careers on the science of Positive Psychology, the study of what makes people flourish. Early discoveries led to hundreds of research studies focused on successful individuals and teams across a broad spectrum of industries and job types. In particular, research on successful learning and workplace environments led to numerous studies of successful teachers and managers. This work included extensive research on individual differences and the environments that best facilitate success. The group of researchers discovered early in their research that

simply measuring the satisfaction of employees was insufficient to create sustainable change. Satisfaction needed to be specified, in terms of its most important elements, and it needed to be measured and reported in a way that could be used by the people who could take action and create change. Further research revealed that change happens most efficiently at a local level, at the level of the front-line, manager-led team. For an executive, the front-line team is his or her direct reports, and for a plant manager, the front-line team is the people he or she manages each day. Studying great managers, Gallup scientists learned that optimal decision-making happens when information regarding decisions is collected at a local level, close to the everyday action.

Dr. Clifton's work merged with Dr. Gallup's work in 1988, when the two organizations combined, enabling the blending of progressive management science with top survey and polling science. Dr. Gallup and Dr. Clifton spent much of their lives studying people. To study people, they wrote questions, recorded the responses, and studied which questions elicited responses that differentiate people and relate to meaningful outcomes. In the case of survey research, some questions are unbiased and elicit meaningful opinions, while others do not. In the case of management research, some questions elicit responses that predict future performance, while others do not. Developing the right questions is an iterative process, in which scientists write questions, and analysis is conducted; the research and questions are refined and rephrased; additional analysis is conducted; the questions are refined and rephrased again; and the process is repeated. The Gallup Organization has followed the iterative process in devising the survey tool that is the subject of this report, Gallup's Q<sup>12</sup> instrument, designed to measure employee engagement.

This section will provide an overview of the many decades of research that have gone into development and validation of Gallup's Q<sup>12</sup> employee engagement instrument. Following this overview, we present a meta-analysis of 166 research studies, exploring the relationship between employee engagement and performance across 125 organizations and 23,910 business or work units.

## **Development of the Q<sup>12</sup>**

---

Beginning in the 1950s, Dr. Clifton began studying work and learning environments in order to determine the factors that contribute positively to those environments and that enable people to capitalize on their unique talents. It was through this early work that Dr. Clifton began using science and the study of strengths to study individuals' frames of reference and attitudes.

From the 1950s to the 1970s, Dr. Clifton continued his research of students, counselors, managers, teachers, and employees. He used various rating scales and interview techniques to study individual differences, analyzing questions and factors that explain differences in people. Concepts studied included "focusing on strengths versus weaknesses," "relationships," "personnel support,"

“friendships,” and “learning.” Various questions were written and tested, including many early versions of the Q<sup>12</sup> items. Ongoing Feedback Techniques were first developed, with the intent of asking questions, collecting data, and encouraging ongoing discussion of the results to provide feedback and potential improvement — a measurement-based feedback process. Exit interviews were also conducted with employees who left organizations, in order to learn causes of employee turnover. A common reason centered on the quality of the manager.

In the 1980s, Gallup scientists continued the iterative process by studying high-performing individuals and teams. Studies involved assessments of individual talents and workplace attitudes. As a starting point for questionnaire design, numerous qualitative analyses were conducted, including interviews and focus groups. Gallup researchers asked top-performing individuals or teams to describe their work environments, and thoughts, feelings, and behaviors related to success. The researchers used qualitative data to generate hypotheses and insights into the distinguishing factors leading to success. From these hypotheses, they wrote and tested questions. They also conducted numerous quantitative studies throughout the 1980s, including exit interviews, to continue to learn causes of employee turnover. Qualitative analyses such as focus groups and interviews formed the basis for lengthy and comprehensive employee surveys, called “Organizational Development Audits” or “Managing Attitudes for Excellence.” Many of these surveys included 100 to 200 items. Quantitative analyses included factor analyses to assess the dimensionality of the survey data, regression analyses to identify uniqueness and redundancies in the data, and criterion-related validity analyses to identify questions that correlate with meaningful outcomes such as overall satisfaction, commitment, and productivity. The scientists developed feedback protocols to facilitate the feedback of survey results to managers and employees. Such protocols, and their use in practice, helped researchers learn which items were most useful in creating dialogue and stimulating change.

One outgrowth of a management research practice focused on both talent and environment was the theory of talent maximization within an organization:

$$\text{Per-person productivity} = \text{Talent} \times (\text{Relationship} + \text{Right Expectation} + \text{Recognition/Reward})$$

These concepts would later become imbedded in the foundational elements of Q<sup>12</sup>.

Over time, SRI and Gallup researchers conducted numerous studies of manager success patterns, focused both on the talents of the manager and the environments that best facilitated success. By integrating knowledge of managerial talent with survey data on employee attitudes, scientists had a unique perspective on what it takes to build a successful workplace environment. Themes such as “individualized perception,” “performance orientation,” “mission,” “recognition,” “learning and growing,” “expectations,” and “the right fit” continued to emerge. In addition to studies of management, researchers conducted numerous studies with successful teachers, students, and learning environments.

In the 1990s, the iterative process continued. During this time, Gallup researchers developed the first version of the Q<sup>12</sup> (“The Gallup Workplace Audit” or GWA), in an effort to efficiently capture the most important workplace information. Both qualitative and quantitative analyses continued. More than 1,000 focus groups were conducted in the decade, and hundreds of instruments were developed, many of them with many additional items. Scientists also continued to use exit interviews; these revealed the importance of the manager in retaining employees. Studies of Q<sup>12</sup> and other survey items were conducted in various countries throughout the world, including the United States, Canada, Mexico, Great Britain, Japan, and Germany. Gallup researchers obtained international cross-cultural feedback on Gallup’s core items, which provided context on the applicability of the items across different cultures. Various scale types were also tested, including variations of 5-point and dichotomous response options.

Quantitative analyses of survey data included descriptive statistics, factor analyses, discriminant analyses, criterion-related validity analyses, reliability analyses, regression analyses, and other correlational analyses. Gallup scientists continued to study the core concepts that differentiated successful from less successful work units, and the expressions that best captured those concepts. In 1997, the criterion-related studies were combined into a meta-analysis to study the relationship of employee satisfaction and engagement (as measured by Q<sup>12</sup>) to business or work unit profitability, productivity, employee retention, and customer satisfaction/loyalty across 1,135 business units (Harter & Creglow, 1997). Meta-analysis also enabled researchers to study the generalizability of the relationship between engagement and outcomes. Results of this confirmatory analysis revealed substantial criterion-related validity for each of the Q<sup>12</sup> items. As criterion-related validity studies are ongoing, the meta-analysis was updated in 1998 (Harter & Creglow, 1998) and included 2,528 business units; in 2000 (Harter & Schmidt, 2000), when it included 7,939 business units; in 2002 (Harter & Schmidt, 2002), when it included 10,885 business units; and in 2003 (Harter, Schmidt, & Killham, 2003), when it included 13,751 business units. This report provides the sixth published iteration of Gallup’s Q<sup>12</sup> meta-analysis, focusing on the relationship between employee engagement and performance. This report expands the number of business units and outcomes studied. The previous meta-analysis examined the relationship between employee engagement and customer loyalty, profitability, productivity, employee turnover, and accidents. We now include absenteeism and merchandise shrinkage (theft and lost merchandise).

Since its final wording and order were completed in 1998, Q<sup>12</sup> has been administered to more than 7 million employees in 112 different countries.

## Introduction to the Study

---

The quality of an organization’s human resources is perhaps the leading indicator of its growth and sustainability. The attainment of a workplace with high-caliber employees starts with the selection of the right people for the right jobs.

Numerous studies have documented the utility of valid selection instruments and systems in the selection of the right people (Schmidt, Hunter, McKenzie, & Muldrow, 1979; Hunter & Schmidt, 1983; Huselid, 1995; Schmidt & Rader, 1999; Harter, Hayes, & Schmidt, 2004). After employees have been selected, they make decisions and take actions every day that can affect the success of their organizations. Many of these decisions and actions are influenced by their own internal motivations and drives. One can also hypothesize that the way employees are treated and the way they treat one another can positively affect their actions — or can place their organizations at risk. For example, researchers have found positive relationships between general workplace attitudes and service intentions, customer perceptions (Schmit & Allscheid, 1995), and individual performance outcomes (Iaffaldano & Muchinsky, 1985). An updated meta-analysis has revealed a substantial relationship between individual job satisfaction and individual performance (Judge, Thoresen, Bono, & Patton, 2001). To date, the vast majority of job satisfaction research and subsequent meta-analyses have collected and studied data at the individual employee level.

There is also evidence at the workgroup or business-unit level that employee attitudes relate to various organizational outcomes. Organizational-level research has focused primarily upon cross-sectional studies. Independent studies found relationships between employee attitudes and performance outcomes such as safety (Zohar, 1980, 2000), customer experiences (Schneider, Parkington, & Buxton, 1980; Ulrich, Halbrook, Meder, Stuchlik, & Thorpe, 1991; Schneider & Bowen, 1993; Schneider, Ashworth, Higgs, & Carr, 1996; Schmit & Allscheid, 1995; Reynierse & Harker, 1992; Johnson, 1996; Wiley, 1991), financials (Denison, 1990; Schneider, 1991), and employee turnover (Ostroff, 1992). A recent study by Batt (2002) used multivariate analysis to examine the relationship between human resource practices (including employee participation in decision making) and sales growth. Gallup has conducted large-scale meta-analyses, most recently studying 13,751 business and work units regarding the concurrent and predictive relationship of employee attitudes (satisfaction and engagement) with safety, customer attitudes, financials, and employee retention (Harter et al., 2003; Harter, Schmidt, & Hayes, 2002; Harter & Schmidt, 2002; Harter & Schmidt, 2000; Harter & Creglow, 1998; Harter & Creglow, 1997). The above studies have found, rather consistently, that there are positive concurrent and predictive relationships between employee attitudes and various important business outcomes.

Even though it has been much more common to study employee opinion data at the individual level, studying data at the business-unit or workgroup level is critical, because that is where the data are typically reported (due to anonymity concerns, employee surveys are reported at a broader business-unit or workgroup level). In addition, business-unit-level research usually provides opportunities to establish links to outcomes that are directly relevant to most businesses — outcomes like customer loyalty, profitability, productivity, employee turnover, and safety variables that are often aggregated and reported at the business-unit level. Another advantage to reporting and studying data at the business-unit or workgroup level is that instrument item scores are of similar reliability to dimension scores for individual-level analysis. This is because at the business-unit or workgroup level, each item score is an average of many individuals'



scores. This means employee surveys reported at a business-unit or workgroup level can be more efficient, i.e., less dependent on length because item-level measurement error is less of a concern. See Harter and Schmidt (2006) for a more complete discussion of job satisfaction research and the advantages obtained by conducting unit-level analyses.

One potential problem with such business-unit-level studies is limited data, due to a limited number of business units (the number of business units becomes the sample size), or limited access to outcome measures that one can compare across business units. For this reason, many of these studies are limited in statistical power, and as such, results from individual studies may appear to conflict with one another. Meta-analysis techniques provide the opportunity to pool such studies together to get more precise estimates of the strength of effects and their generalizability.

This paper's purpose is to present the results of an updated meta-analysis of the relationship between employee workplace perceptions and business-unit outcomes, based on currently available data collected with Gallup clients. The focus of this study is on Gallup's Q<sup>12</sup> instrument. The Q<sup>12</sup> items — which were selected because of their importance at the business-unit or workgroup level — measure employee perceptions of the quality of people-related management practices in their business units.

## Description of the Q<sup>12</sup>

---

In short, the development of the GWA (Q<sup>12</sup>) was based on over 30 years of accumulated quantitative and qualitative research. Its reliability, convergent validity, and criterion-related validity have been extensively studied. It is an instrument validated through prior psychometric studies as well as practical considerations regarding its usefulness for managers in creating change in the workplace.

In designing the items included in the Q<sup>12</sup>, researchers took into account that, from an actionability standpoint, there are two broad categories of employee survey items: those that measure attitudinal outcomes (satisfaction, loyalty, pride, customer service intent, and intent to stay with the company) and those that measure actionable issues that drive the above outcomes. The Q<sup>12</sup> measures the actionable issues for management — those predictive of attitudinal outcomes such as satisfaction, loyalty, pride, and so on. On Gallup's standard Q<sup>12</sup> instrument, following an overall satisfaction item are 12 items measuring issues we have found to be actionable at the supervisor or manager level in the company — items measuring the extent to which employees are “engaged” in their work.

**The Q<sup>12</sup> statements are as follows:**

- Q00. (Overall Satisfaction) On a five-point scale, where “5” is *extremely satisfied* and “1” is *extremely dissatisfied*, how satisfied are you with (your company) as a place to work?
- Q01. I know what is expected of me at work.
- Q02. I have the materials and equipment I need to do my work right.
- Q03. At work, I have the opportunity to do what I do best every day.
- Q04. In the last seven days, I have received recognition or praise for doing good work.
- Q05. My supervisor, or someone at work, seems to care about me as a person.
- Q06. There is someone at work who encourages my development.
- Q07. At work, my opinions seem to count.
- Q08. The mission or purpose of my company makes me feel my job is important.
- Q09. My associates or fellow employees are committed to doing quality work.
- Q10. I have a best friend at work.
- Q11. In the last six months, someone at work has talked to me about my progress.
- Q12. This last year, I have had opportunities at work to learn and grow.

*These statements (Q00-Q12) are proprietary and copyrighted by The Gallup Organization. They cannot be reprinted or reproduced in any manner without the written consent of The Gallup Organization. Copyright © 1992-1999, The Gallup Organization, Princeton, NJ. All rights reserved.*

The current standard is to ask each employee to rate the above statements (a census survey — median participation rate is 85%) using six response options (from 5=strongly agree to 1=strongly disagree; the sixth response option — don’t know/does not apply — is unscored). Because it is a satisfaction item, the first item is scored on a satisfaction scale rather than on an agreement scale. Regression analyses (Harter, Schmidt, & Hayes, 2002) indicate that employee engagement accounts for nearly all of the performance-related variance (composite performance) accounted for by the overall satisfaction measure.

Therefore, the focus of this report is on employee engagement (as measured by statements Q01-Q12 below).

The reader will notice that, while these items measure issues that the manager or supervisor can influence, only one item contains the word “supervisor.” This is because it is realistic to assume that numerous people in the workplace can influence whether someone’s expectations are clear, whether he or she feels cared about, and so on. The manager’s or supervisor’s position, though, allows him or her to take the lead in establishing a culture that values behaviors that support these perceptions. The following is a brief discussion of the conceptual relevancy of each of the 13 items:

- Q00. *Overall satisfaction.* The first item on the survey measures an overall attitudinal outcome: satisfaction with one’s company. One could argue that in and of itself, it is difficult to act on the results of this item. Other issues, like those measured in the following 12 items, explain why people are satisfied, and why they become engaged and affect outcomes.
- Q01. *Expectations.* Defining and clarifying the outcomes that are to be achieved is perhaps the most basic of all employee needs and manager responsibilities. How these outcomes are defined and acted upon will vary from business unit to business unit, depending on the goals of the business unit.
- Q02. *Materials and equipment.* Getting people what they need to do their work is important in maximizing efficiency, in demonstrating to employees that their work is valued, and in showing that the company is supporting them in what they are asked to do. Great managers keep this perception objective by helping employees see how their requests for materials and equipment connect to important outcomes.
- Q03. *Opportunity to do what I do best.* Helping people get into roles where they can most fully use their inherent talents is the ongoing work of great managers. Learning about individual differences through experience and assessment can help the manager position people efficiently, within and across roles.
- Q04. *Recognition for good work.* When managers ask employees who are performing at a high level whether they are suffering from too much recognition, they rarely, if ever, get an affirmative response. Another ongoing management challenge is to understand how each person prefers to be recognized, to make it objective and real by basing it on performance, and to do it frequently.
- Q05. *Someone at work cares about me.* For each person, feeling “cared about” may mean something different. The best managers listen to individuals, and respond to their unique needs. In addition, they find

the connection between the needs of the individual and the needs of the organization.

- Q06. *Encourages my development.* How employees are coached can influence how they perceive their future. If the manager is helping the employee improve as an individual by providing opportunities that are in sync with the employee's talents, both the employee and the company will profit.
- Q07. *Opinions count.* Asking for the employee's input, and considering that input as decisions are made, can often lead to better decisions. This is because employees are often closer than the manager is to individuals and variables that affect the overall system. In addition, when employees feel they are involved in decisions, they take greater ownership of the outcomes.
- Q08. *Mission/Purpose.* Great managers often help people see not only the purpose of their work, but also how each person's work influences and relates to the purpose of the organization and its outcomes. Reminding employees of the big-picture impact of what they do each day is important, whether it is how their work influences the customer, safety, or the public.
- Q09. *Associates committed to quality.* Managers can influence the extent to which employees respect one another by selecting conscientious employees, providing some common goals and metrics around quality, and increasing associates' frequency of opportunity for interaction.
- Q10. *Best friend.* Managers vary in the extent to which they create opportunities for people at work to get to know one another, and in whether they value close, trusting relationships at work. The best managers do not subscribe to the idea that there should be no close friendships at work; instead, they free people to get to know one another, which is a basic human need. This, then, can influence communication, trust, and other outcomes.
- Q11. *Progress.* Providing a structured time to discuss each employee's progress, achievements, goals, and so on, is important for both managers and employees. Great managers regularly meet with individuals, both to learn from them and to give them guidance. This give-and-take helps both managers and employees make better decisions.
- Q12. *Learn and grow.* In addition to having a need to be recognized for good work, most employees have a need to know they are improving and have chances to improve themselves. Great managers pick training that will benefit the individual and the organization.

As a total instrument (sum or mean of items 01-12), the Q<sup>12</sup> has a Cronbach's alpha of .91 at the business-unit level. The meta-analytic convergent validity of the equally weighted mean (or sum) of items 01-12 (GrandMean) to the equally weighted mean (or sum) of additional items in longer surveys (measuring all known facets of job satisfaction and engagement) is .91. This provides evidence that the Q<sup>12</sup>, as a composite measure, captures the general factor in longer employee surveys. Individual items correlate to their broader dimension true-score values, on average, at .69.

As mentioned, this is the sixth published iteration of Q<sup>12</sup> business-unit-level meta-analysis. The current meta-analysis includes a larger number of studies, business units, and industries represented, and two additional outcomes studied (absenteeism and shrinkage). This meta-analysis includes a much larger number of studies with safety as a dependent variable, and more studies from countries outside the United States (20 studies conducted exclusively outside the U.S., including data from Asia, Canada, Central America, Europe, and South America). This meta-analysis also includes updated estimates of reliabilities across business units and includes all available Gallup studies (whether published or unpublished) and has no risk of publication bias.

# Q<sup>12</sup>® Meta-Analysis

*Meta-Analysis, Hypothesis, Methods, and Results*

## Meta-Analysis

---

A meta-analysis is a statistical integration of data accumulated across many different studies. As such, it provides uniquely powerful information, because it controls for measurement and sampling errors and other idiosyncrasies that distort the results of individual studies. A meta-analysis eliminates biases and provides an estimate of true validity or true relationship between two or more variables. Statistics typically calculated during meta-analyses also allow the researcher to explore the presence, or lack thereof, of moderators of relationships. More than 1,000 meta-analyses have been conducted in the psychological, educational, behavioral, medical, and personnel selection fields. The research literature in the behavioral and social sciences fields includes a multitude of individual studies with apparently conflicting conclusions. Meta-analysis, however, allows the researcher to estimate the mean relationship between variables and make corrections for artifactual sources of variation in findings across studies. It provides a method by which researchers can determine whether validities and relationships generalize across various situations (e.g., across firms or geographical locations).

This paper will not provide a full review of meta-analysis. Rather, the authors encourage readers to consult the following sources for both background information and detailed descriptions of the more recent meta-analytic methods: Schmidt (1992); Hunter and Schmidt (1990, 2004); Lipsey and Wilson (1993); Bangert-Drowns (1986); and Schmidt, Hunter, Pearlman, and Rothstein-Hirsh (1985).

## Hypothesis and Study Characteristics

---

**The hypotheses examined for this meta-analysis were as follows:**

*Hypothesis 1:* Business-unit-level employee engagement will have positive average correlations with the business-unit outcomes of customer loyalty, productivity, and profitability, and negative correlations with employee turnover, employee safety incidents (accidents), absenteeism, and shrinkage (theft).

*Hypothesis 2:* The correlations between engagement and business-unit outcomes will generalize across organizations for all business-unit outcomes. That is, these

correlations will not vary substantially across organizations, and in particular, there will be few, if any, organizations with zero or negative correlations.

A total of one hundred sixty-six (166) studies for 125 independent organizations are included in Gallup's inferential database — studies conducted as proprietary research for various organizations. In each Q<sup>12</sup>, one or more of the Q<sup>12</sup> items were used (as a part of standard policy, starting in 1997, all items were included in all studies), and data were aggregated at the business-unit level and correlated with the following aggregate business-unit performance measures:

- Customer metrics (referred to as customer loyalty)
- Profitability
- Productivity
- Turnover
- Safety & Health
- Absenteeism
- Shrinkage

That is, in these analyses, the unit of analysis was the business unit, not the individual employee.

Pearson correlations were calculated, estimating the relationship of business-unit average measures of employee engagement (the mean of the Q<sup>12</sup> items) to each of these seven general business outcomes. Correlations were calculated across business units within each company, and these correlation coefficients were entered into a database. The researchers then calculated mean validities, standard deviations of validities, and validity generalization statistics for each item for each of the seven business-unit outcome measures.

As with previous meta-analyses, some of the studies were concurrent validity studies, where engagement and performance were measured within roughly the same time period, or with engagement measurement slightly trailing the performance measurement (since engagement is relatively stable and a summation of the recent past, such studies are considered “concurrent”). Predictive validity studies involve measuring engagement at time 1 and performance at time 2. “Predictive” validity estimates were obtained for approximately 43% of the companies included in this meta-analysis. This paper does not directly address issues of causality, which are best addressed with meta-analytic longitudinal data, consideration of multiple variables, and path analysis. Issues of causality are discussed and examined extensively in other sources (Harter, Schmidt, Asplund, & Killham, 2005).

Studies for the current meta-analysis were selected so that each company was represented once in each analysis. For several companies, multiple studies were conducted. In order to include the best possible information for each company represented in the study, some basic rules were used. If two concurrent studies were conducted for the same client (where Q<sup>12</sup> and outcome data were collected

concurrently, i.e., in the same year), then the weighted average effect sizes across the multiple studies were entered as the value for that company. If a company had both a concurrent and a predictive study (where the Q<sup>12</sup> was collected in year 1 and outcomes were tracked in year 2), then the effect sizes from the predictive study were entered. If a company had multiple predictive studies, then the mean of the correlations in these studies was entered. If sample sizes varied substantially in repeated studies for an organization, the study with the larger of the sample sizes was used.

- For 54 companies, there were studies that examined the relationship between business-unit employee perceptions and customer perceptions. Customer perceptions included customer metrics, patient metrics, and student ratings of teachers. These metrics included measures of loyalty, satisfaction, and engagement. The largest representation of studies included loyalty metrics (i.e., likelihood to recommend or repeat business), so we refer to customer metrics as customer loyalty in this study. Instruments varied from study to study. The general index of customer loyalty was an average score of the items included in each measure. A growing number of studies include “customer engagement” as the metric of choice, which measures the emotional connection between the customers and the organization that serves them. For more information on the interaction of employee and customer engagement, see Fleming, Coffman, and Harter (2005), and Harter, Asplund, and Fleming (2004).
- Profitability studies were available for 63 companies. Definition of profitability typically was a percentage profit of revenue (sales). In several companies, the researchers used — as the best measure of profit — a difference score from the prior year or a difference from a budgeted amount, because it represented a more accurate measure of each unit’s relative performance. As such, a control for opportunity was used when profitability figures were deemed less comparable from one unit to the next. For example, a difference variable involved dividing profit by revenue for a business unit and then subtracting a budgeted percentage from this percentage. In every case, profitability variables were measures of margin, and productivity variables (which follow) were measures of amount produced.
- Productivity studies were available for 83 companies. Measures of business-unit productivity consisted of one of the following: financials (i.e., revenue/sales dollars per person or patient), quality (i.e., managerial evaluation of all available productivity measures), quantity produced, or student achievement scores. In a few cases, this was a dichotomous variable (top-performing business units = 2; less successful units = 1). As with profitability, in many cases it was necessary for the researchers to control the financial metrics for opportunity by comparing results to a performance goal or prior year figure.



- Turnover data were available for 54 companies. The turnover measure was the annualized percentage of employee turnover for each business unit.
- Safety & Health data were available for 27 companies. Safety and health measures included lost workday/time incident rate, percentage of workdays lost due to incidents or worker's compensation claims, number of incidents, or incident rates. One large healthcare organization provided risk-adjusted mortality rates and risk-adjusted complication rates for each hospital.
- Absenteeism data were included for 6 companies. Absenteeism measures included the average number of days missed per person for each work unit divided by the total days available for work. Three companies provided sick days or hours and three provided total absenteeism rates.
- Measures of shrinkage were provided by 3 companies. Shrinkage is defined as the dollar amount of unaccounted-for lost merchandise, which could be due to employee theft, customer theft, or lost merchandise. Given the varying size of locations, shrinkage was calculated as a percentage of total revenue, or a difference from an expected target.
- The overall study involved 681,799 independent employee responses to surveys and 23,910 independent business units in 125 companies, an average of 29 employees per business unit and 191 business units per company. One hundred sixty-six (166) research studies were conducted across the 125 companies.
- Table 1, which follows, provides a summary of studies (per company) sorted by industry type. It is evident that there is considerable variation in the industry types represented, as companies from 37 industries provided studies. Each of the general government industry classifications (via SIC codes) is represented, with the largest number of companies represented in services, retail, and manufacturing industries. The largest number of business units is in retail, financial, and transportation/public utilities. Of the specific industry classifications, finance-depository, communications, retail-food, and services-health are most frequently represented.

Table 1 — Summary of Studies by Industry

<i>Industry Type</i>	Number of		
	Companies	Business Units	Respondents
<b>Finance</b>			
Credit	2	59	581
Depository	10	4,787	62,291
Insurance	2	174	7215
Mortgage	1	27	985
Nondepository	1	94	2,038
Security	3	86	2,785
<b>Manufacturing</b>			
Building Materials	1	8	1,335
Consumer Goods	2	146	5,797
Food	4	161	20,062
Industrial Equipment	1	89	639
Instrument	7	81	2,169
Paper	1	60	17,243
Pharmaceutical	2	124	992
Printing	2	35	716
<b>Materials and Construction</b>	2	218	19,926
<b>Real Estate</b>	2	185	3,790
<b>Retail</b>			
Automotive	2	105	8,313
Building Materials	2	793	43,763
Clothes	3	279	16,795
Department Stores	1	567	7,362
Eating	6	732	37,191
Electronics	4	960	96,287
Entertainment	1	106	1,051
Food	4	3,745	73,630
Industrial Equipment	1	11	484
Miscellaneous	8	2,599	76,439
Pharmaceutical	1	180	3,004
— continued —			

Table 1 — Summary of Studies by Industry (cont.)

<i>Industry Type</i>	<b>Number of</b>		
	<b>Companies</b>	<b>Business Units</b>	<b>Respondents</b>
<b>Services</b>			
Business	1	20	600
Education	7	310	10,746
Government	4	240	8,336
Health	23	2,218	80,990
Hospitality	1	30	2,612
Hotels	4	202	10,924
Recreation	1	14	288
<b>Transport/Public Utilities</b>			
Trucking	1	96	6,213
Communications	5	4,138	43,633
Electric	2	231	4,574
<b>Total</b>			
Financial	19	5,227	75,895
Manufacturing	20	704	48,953
Materials & Construction	2	218	19,926
Real Estate	2	185	3,790
Retail	33	10,077	364,299
Services	41	3,034	114,496
Transportation/Public Utilities	8	4,465	54,420
<b>Total</b>	<b>125</b>	<b>23,910</b>	<b>681,799</b>

Table 2 provides a summary of studies (per company) sorted by business or operational unit type. There is also considerable variation in type of business unit, ranging from stores to plants/mills to departments to schools. Overall, 19 different types of business units are represented; the largest number of companies had studies of workgroups, stores, or bank branches. Likewise, workgroups, stores, and bank branches have the highest proportional representation of business/operating units.

**Table 2 — Summary of Business/Operating Unit Types**

<i><b>Business/Operating Unit Type</b></i>	<b>Number of</b>		
	<b>Companies</b>	<b>Business Units</b>	<b>Respondents</b>
Bank Branch	13	5,038	65,653
Call Center	2	41	4,152
Call Center Department	4	120	2,409
Cost Center	2	68	10,597
Dealership	2	105	8,313
Division	2	29	5,090
Hospital	4	248	31,167
Hotel	3	101	10,111
Location	7	130	6,730
Mall	2	185	3,790
Patient Care Unit	2	177	4,433
Plant/Mill	4	131	38,155
Region	1	96	6,213
Restaurant	5	369	21,183
Sales Division	4	34	1,515
Sales Team	4	313	16,791
School	6	296	10,496
Store	24	9,097	327,670
Workgroup	34	7,332	107,331
<b>Total</b>	<b>125</b>	<b>23,910</b>	<b>681,799</b>

## Meta-Analytic Methods Used

---

Analyses included weighted average estimates of true validity, estimates of standard deviation of validities, and corrections made for sampling error, measurement error in the dependent variables, and range variation and restriction in the independent variable (Q<sup>12</sup> GrandMean) for these validities. An additional analysis was conducted, correcting for independent-variable measurement error. The most basic form of meta-analysis corrects variance estimates only for sampling error. Other corrections recommended by Hunter and Schmidt (1990, 2004) include correction for measurement and statistical artifacts, such as range restriction and measurement error in the performance variables gathered. The definitions of the above procedures are provided in the sections that follow.

Gallup researchers gathered performance-variable data for multiple time periods to calculate the reliabilities of the business performance measures. Since these multiple measures were not available for each study, the researchers used artifact distributions meta-analysis methods (Hunter & Schmidt, 1990, pp. 158-197) to correct for measurement error in the performance variables. The artifact distributions developed were based on test-retest reliabilities, where they were available, from various studies. The procedure followed for calculation of business-unit outcome-measure reliabilities was consistent with Scenario 23 in Schmidt and Hunter (1996). To take into account that some change in outcomes (stability) is a function of real change, test-retest reliabilities were calculated using the following formula:

$$(r_{12} \times r_{23})/r_{13}$$

*Where r<sub>12</sub> is the correlation of the outcome measured at time 1 with the same outcome measured at time 2; r<sub>23</sub> is the correlation of the outcome measured at time 2 with the outcome measured at time 3; and r<sub>13</sub> is the correlation of the outcome measured at time 1 with the outcome measured at time 3.*

The above formula factors out real change (which is more likely to occur from time period 1-3 than from time periods 1-2 or 2-3) from random changes in business-unit results caused by measurement error, data collection errors, sampling errors (primarily in customer measures), and uncontrollable fluctuations in outcome measures. Some estimates were available for quarterly data, some for semiannual data, and others for annual data. The average time period in artifact distributions used for this meta-analysis was equal to the average time period across studies for each criterion type. See Appendix A for a listing of the reliabilities used in the corrections for measurement error. Artifact distributions for reliability were collected for customer measures, profitability, productivity, turnover, and safety. But they were not collected for absenteeism and shrinkage, because they were not available at the time of this study. Therefore, the assumed reliability for absenteeism and shrinkage was 1.00, resulting in downwardly biased true validity estimates (the estimates of validity reported here are lower than reality). Artifact distributions for the last two variables will be added to upcoming reports as they become available.

It could be argued that, because the independent variable (employee engagement as measured by Q<sup>12</sup>) is used in practice to predict outcomes, the practitioner must live with the reliability of the instrument he or she is using. However, correcting for measurement error in the independent variable answers the theoretical question of how the actual constructs (true scores) relate to each other. Therefore, we present analyses both before and after correcting for independent variable reliability. Appendix B presents the distributions of reliabilities for the GrandMean of Q<sup>12</sup>. These values were computed in the same manner as were those for the business-unit outcomes.

In correcting for range variation and range restriction, there are fundamental, theoretical questions that need to be considered relating to whether such correction is necessary. In personnel selection, validities are routinely corrected for range restriction because, in selecting applicants for jobs, those scoring highest on the predictor are typically selected. This results in explicit range restriction that biases observed correlations downward (i.e., attenuation). In the employee satisfaction and engagement arena, one could argue that there is no explicit range restriction because we are studying results as they exist in the workplace. Work units are not selected based on scores on the predictor (Q<sup>12</sup> scores). However, in studying companies, we have observed that there is variation across companies in standard deviations of indices. One hypothesis for why this variation occurs is that companies vary in how they encourage employee satisfaction and engagement initiatives and in how they have or have not developed a common set of values and a common culture. Therefore, the standard deviation of the population of business units across organizations studied will be greater than the standard deviation within the typical company. This variation in standard deviations across companies can be thought of as indirect range restriction (as opposed to direct range restriction). Improved indirect range-restriction corrections have been incorporated into this meta-analysis (Hunter, Schmidt, & Le, in press).

Since the development of the Q<sup>12</sup>, Gallup has collected descriptive data on more than 7 million respondents, 600,000 business units or workgroups, and 500 companies. This accumulation of data indicates that the standard deviation within a given company is approximately three-fourths the standard deviation in the population of all business units. In addition, the ratio of standard deviation for a given company relative to the population value varies from company to company. Therefore, if one goal is to estimate the effect size in the population of all business units (arguably a theoretically important issue), then correction should be made based on such available data. In the observed data, correlations are attenuated for companies with less variability across business units than the population average, and vice versa. As such, variability in standard deviations across companies will create variability in observed correlations and is therefore an artifact that can be corrected for in interpreting the generalizability of validities. Appendices in Harter and Schmidt (2000) provided artifact distributions for range-restriction/variation corrections used for meta-analysis. These artifact distributions have since been updated substantially. Due to the increased size of these tables, they are not included in this report. They resemble those reported in the earlier study, but with a larger number of entries. The

following excerpt provides an overview of meta-analysis conducted using artifact distributions:

In any given meta-analysis, there may be several artifacts for which artifact information is only sporadically available. For example, suppose measurement error and range restriction are the only relevant artifacts beyond sampling error. In such a case, the typical artifact distribution-based meta-analysis is conducted in three stages:

- First, information is compiled on four distributions: the distribution of the observed correlations, the distribution of the reliability of the independent variable, the distribution of the reliability of the dependent variable, and the distribution of the range departure. There are then four means and four variances compiled from the set of studies, with each study providing whatever information it contains.
- Second, the distribution of observed correlations is corrected for sampling error.
- Third, the distribution corrected for sampling error is then corrected for error of measurement and range variation (Hunter & Schmidt, 1990, pp. 158-159).

In this study, statistics are calculated and reported at each level of analysis, starting with the observed correlations and then correcting for sampling error, measurement error, and, finally, range variation. Both within-company range-variation corrections (to correct validity generalization estimates) and between-company range-restriction corrections (to correct for differences in variation across companies) were made. Between-company range-restriction corrections are relevant in understanding how engagement relates to performance across the business units of all companies. As alluded to, we have applied the indirect range-restriction correction procedure to this meta-analysis (Hunter et al., in press).

The meta-analysis includes an estimate of the mean sample-size-weighted validity and the variance across the correlations — again weighting each validity by its sample size. The amount of variance predicted for weighted correlations on the basis of sampling error was also computed. The following is the formula to calculate variance expected from sampling error in "bare bones" meta-analyses, using the Hunter et al. (in press) technique referred to on the previous page:

$$s_e^2 = (1 - \bar{r}^2)^2 / (\bar{N} - 1)$$

Residual standard deviations were calculated by subtracting the amount of variance due to sampling error, the amount of variance due to study differences in measurement error in the dependent variable, and the amount of variance due

to study differences in range variation from the observed variance. To estimate the true validity standard deviations, the residual standard deviation was adjusted for bias due to mean unreliability and mean range restriction. The amount of variance due to sampling error, measurement error, and range variation was divided by the observed variance to calculate the total percentage variance accounted for. One rule of thumb adopted from the literature is that, if over 75% of variance in validities across studies is due to sampling error and other artifacts, the validity is assumed generalizable.

As in Harter, Schmidt, and Hayes (2002) and Harter et al. (2003), we calculated the correlation of engagement to composite performance. This calculation assumes managers are managing toward multiple outcomes simultaneously and that each outcome occupies some space in the overall evaluation of performance. To calculate the correlation to the composite index of performance, we used the Mosier (1943) formula to determine the reliability of the composite measure (as described in Harter, Schmidt, and Hayes, 2002), with updated reliability distributions and updated intercorrelations of the outcome measures. In addition, given the increase in number of outcomes studied, we included absenteeism and shrinkage as outcomes in the composite performance definition. The reliability of the composite metric is .89. Composite performance was measured as the equally weighted sum of customer loyalty, turnover (reverse scored as retention), safety (accidents reverse scored), absenteeism (reverse scored), shrinkage (reverse scored), and financials (with profitability and productivity equally weighted). We also calculated composite performance as the equally weighted sum of the most direct outcomes of engagement — customer loyalty, turnover (reverse scored as retention), and safety (accidents reverse scored), absenteeism (reverse scored), and shrinkage (reverse scored). The reliability of this composite variable is .87.

In our research, we used the Schmidt and Le (2004) meta-analysis package (the artifact distribution meta-analysis program with correction for indirect range restriction). The program package is described in Hunter and Schmidt (2004).

## Results

---

The focus of analyses for this report is on the relationship between overall employee engagement (defined by an equally weighted GrandMean of Q<sup>12</sup>) and a variety of outcomes. Table 3 provides meta-analytic and validity generalization statistics for the relationship between employee engagement and business performance for each of the seven outcomes studied.

Mean observed correlations and standard deviations are followed by two forms of true validity estimation. The first corrects for range variation within companies and dependent-variable measurement error. This range-restriction correction places all companies on the same basis in terms of variability of employee engagement across business units. These results can be viewed as estimating the relationships across business units within the average company.



The second corrects for range restriction across the population of business units and dependent-variable measurement error. Estimates that include the latter range-restriction correction apply to interpretations of effects in business units across companies, as opposed to effects expected within a given company. Because there is more variation in business units across companies than there is within the average company, effect sizes are higher when true validity estimates are calculated for business units across companies.

For instance, observe the estimates relative to the customer loyalty criteria. Without the between-company range-restriction correction (which is relevant to the effect within the typical company), the true validity value of employee engagement is .22 with a 90% CV of .22. With the between-company range-restriction correction (which is relevant to business units across companies), the true validity value of employee engagement is .33 with a 90% CV of .32. As in prior studies, findings here show high generalizability across companies in the relationship between employee engagement and customer loyalty metrics, profitability, productivity, employee turnover, and safety outcomes. Most of the variability in correlations across companies was due to sampling error in individual studies, and for each of these five outcomes, all of the variability in correlations across companies can be attributed to artifacts (sampling error, range variation, and measurement error). In other words, the true validity is essentially the same for each company studied. For the two new outcomes (absenteeism and shrinkage), results indicate high generalizability across the three companies with shrinkage data, and substantial generalizability for absenteeism. However, not all of the variance in correlations of employee engagement and absenteeism is explained by artifacts. It is possible that this is due to a lack of reliability estimates for this outcome. Once reliability estimates become available and as more studies are added to the meta-analysis, future research may shed light on this. Regardless, 76% of the variability in the correlations across the six companies is explained by the artifacts we were able to account for — substantial evidence of generalizability (Schmidt & Hunter, 1977).

Table 3 — Meta-Analysis of Relationship Between Employee Engagement and Business-Unit Performance

	<b>Customer</b>	<b>Profit-ability</b>	<b>Produc-tivity</b>	<b>Turnover</b>	<b>Safety Inci-dents</b>	<b>Absen-teeism</b>	<b>Shrink-age</b>
Number of Bus. Units	8,541	14,044	15,152	15,871	2,811	598	806
Number of r's	54	63	83	54	27	6	3
Mean Observed r	0.17	0.11	0.15	-0.12	-0.15	-0.21	-0.24
Observed SD	0.09	0.08	0.09	0.06	0.12	0.15	0.03
True Validity r <sup>1</sup>	0.22	0.12	0.16	-0.18	-0.18	-0.21	-0.25
True Validity SD <sup>1</sup>	0.00	0.00	0.00	0.00	0.00	0.08	0.00
True Validity <sup>2</sup>	0.33	0.19	0.24	-0.28	-0.28	-0.32	-0.37
True Validity SD <sup>2</sup>	0.00	0.00	0.00	0.00	0.00	0.11	0.00
% Variance Acct'd for — sampling error	70	67	62	91	71	42	365
% Variance Acct'd for <sup>1</sup>	131	104	109	191	103	75	1402
% Variance Acct'd for <sup>2</sup>	132	105	109	193	104	76	1426
90% CV <sup>1</sup>	0.22	0.12	0.16	-0.18	-0.18	-0.12	-0.25
90% CV <sup>2</sup>	0.32	0.19	0.24	-0.28	-0.28	-0.19	-0.37

SD = Standard Deviation

<sup>1</sup> Includes correction for range variation within companies and dependent-variable measurement error<sup>2</sup> Includes correction for range restriction across population of business units and dependent-variable measurement error

In summary, for the overall measures of engagement shown in Table 3, the strongest effects were found relative to outcomes that are likely to be more direct outcomes of employee engagement (customer loyalty metrics, employee turnover, safety, absenteeism, and shrinkage). Correlations were positive and generalizable relative to profitability and productivity (often defined as sales) criteria, but of slightly lower magnitude. This may be because profitability and other financial variables are influenced indirectly by employee engagement and more directly by the customer, employee turnover, safety, absenteeism, and shrinkage. The next section will explore the practical utility of the observed relationships.

As in Harter, Schmidt, and Hayes (2002), we calculated the correlation of employee engagement to composite performance. As defined earlier, Table 4 provides the correlations and d-values for four analyses: the observed correlations, correction for dependent-variable measurement error, correction for dependent-variable measurement error and range restriction across companies, and correction for dependent-variable measurement error, range restriction, and independent-variable measurement error (true score correlation).

As with previous meta-analyses, the effect sizes presented in Table 4 indicate substantial relationships between engagement and composite performance.

Business units in the top half on engagement within companies have .58 standard deviation units' higher composite performance in comparison to those in the bottom half on engagement.

Across companies, business units in the top half on engagement have .77 standard deviation units' higher composite performance in comparison to those in the bottom half on engagement.

After correcting for all available study artifacts (examining the true score relationship), business units in the top half on employee engagement have .85 standard deviation units' higher composite performance in comparison to those in the bottom half on engagement. This is the true score effect expected over time, across all business units.

Table 4 — Correlation of Employee Engagement to Composite Business-Unit Performance — All Outcomes

Analysis	Correlation of Engagement to Performance
Observed r	0.33
d	0.55
r corrected for dependent-variable measurement error	0.35
d	0.58
r corrected for dependent-variable measurement error and range restriction across companies	0.45
d	0.77
r corrected for dependent-variable measurement error, range restriction across companies, and independent-variable measurement error	0.49
d	0.85

As alluded to, some outcomes are the direct consequence of employee engagement (i.e., employee turnover, customer loyalty, safety, absenteeism, and shrinkage), and other outcomes are a more downstream result of intermediary outcomes (i.e., sales and profit). For this reason, we have also calculated the composite correlation to short-term outcomes. Table 5 again indicates a substantial relationship between engagement and composite performance. Observed correlations and d-values are of approximately the same magnitude as those reported in Table 4, but slightly lower (most likely because the direct outcomes do not occupy all of the performance criterion space).

Business units in the top half on engagement within companies have .55 standard deviation units' higher performance on direct outcomes in comparison to those in the bottom half. Across companies, the difference is .71 standard deviation units. After correcting for all available artifacts, the difference is .81 standard deviation units.

**Table 5 — Correlation of Employee Engagement to Composite Business-Unit Performance — Direct Outcomes  
(Customer, Turnover, Safety, Absenteeism, Shrinkage)**

<b>Analysis</b>	<b>Correlation of Engagement to Performance</b>
Observed r d	0.31 0.51
r corrected for dependent-variable measurement error d	0.33 0.55
r corrected for dependent-variable measurement error and range restriction across companies d	0.42 0.71
r corrected for dependent-variable measurement error, range restriction across companies, and independent- variable measurement error d	0.47 0.81

# Q<sup>12</sup>® Meta-Analysis

## *Utility Analysis: Practicality of the Effects*

### **Utility Analysis**

---

In the past, studies of job satisfaction's relationship to performance have had limited analysis of the utility of the reported relationships. Correlations have often been discounted as trivial without an effort to understand the potential utility, in practice, of the relationships. The Q<sup>12</sup> includes items Gallup researchers have found to be influenceable by the local manager. As such, understanding the practical utility of potential changes is critical.

The research literature includes a great deal of evidence that numerically small or moderate effects often translate into large practical effects (Abelson, 1985; Carver, 1975; Lipsey, 1990; Rosenthal & Rubin, 1982; Sechrest & Yeaton, 1982). As shown in Table 6, this is, in fact, the case here. Effect sizes referenced in this study are consistent with or above other practical effect sizes referenced in other reviews (Lipsey & Wilson, 1993).

A more intuitive method of displaying the practical value of an effect is that of binomial effect size displays, or BESDs (Rosenthal & Rubin, 1982; Grissom, 1994). BESDs typically depict the success rate of a treatment versus a control group as a percentage above the median on the outcome variable of interest.

BESDs can be applied to results of this study. Table 6 provides the percentage of business units above the median on the outcomes of interest for high- and low-scoring business units on the employee engagement (Q<sup>12</sup>) composite measure. True validity estimates (correcting for measurement error only in the dependent variable) were used for analysis of business units **both** within companies and across companies.

One can see from Table 6 that there are meaningful differences between the top and bottom halves; again, the top half is defined as the average of business units scoring in the highest 50% on the Q<sup>12</sup>, and business units scoring in the bottom half comprise the lowest 50%. It is clear from Table 6 that management would learn a great deal more about success if it studied what was going on within top-half business units rather than bottom-half units. Within companies, business units in the top half on employee engagement had, on average, a 56% higher success rate on customer loyalty metrics [i.e.,  $(61\% - 39\%) \div 39\% = 56.4\%$ ], a 44% higher success rate on turnover (lower probability of turnover), a 38% higher success rate on productivity outcomes, and a 27% higher success rate on profitability. For the safety variable, business units in the top half on employee engagement had, on average, a 44%

higher success rate (lower probability of injuries or lost workdays). For absenteeism, business units in the top half on employee engagement had, on average, a 56% higher success rate (lower probability of high absenteeism). For shrinkage, business units in the top half on employee engagement had, on average, a 70% higher success rate (lower probability of high shrinkage). For business units across companies, those in the top half on employee engagement had, on average, a 103% higher success rate on customer metrics, a 78% higher success rate on turnover (lower probability of turnover), a 63% higher success rate on productivity outcomes, a 50% higher success rate on profitability outcomes, a 78% higher success rate on safety (lower probability of injuries or lost workdays), a 94% higher success rate on absenteeism (lower probability of high absenteeism), and a 123% higher success rate on shrinkage (lower probability of high shrinkage).

With regard to composite business-unit performance, business units in the top half on employee engagement have a 113% higher success rate within their own company, and a 170% higher success rate across business units in all companies studied. In other words, business units high in employee engagement more than double their odds of above-average composite performance within their own companies, and nearly triple their chances for above-average success across business units in all companies.

Table 6 — BESDs for Employee Engagement and Outcomes

<b>Employee Engagement GrandMean</b>			
<b>Business Units Within Companies</b>		<b>Business Units Across Companies</b>	
	% Above Median on Customer Metrics		% Above Median on Customer Metrics
Top Half	61	Top Half	67
Bottom Half	39	Bottom Half	33
	% Above Median on Profitability Metrics		% Above Median on Profitability Metrics
Top Half	56	Top Half	60
Bottom Half	44	Bottom Half	40
	% Above Median on Productivity Metrics		% Above Median on Productivity Metrics
Top Half	58	Top Half	62
Bottom Half	42	Bottom Half	38
	% Below Median on Turnover Metrics		% Below Median on Turnover Metrics
Top Half	59	Top Half	64
Bottom Half	41	Bottom Half	36
	% Below Median on Safety Incidents		% Below Median on Safety Incidents
Top Half	59	Top Half	64
Bottom Half	41	Bottom Half	36
— continued —			



Table 6 — BESDs for Employee Engagement and Outcomes (cont.)

<b>Employee Engagement GrandMean</b>			
<b>Business Units Within Companies</b>		<b>Business Units Across Companies</b>	
	% Below Median on Absenteeism		% Below Median on Absenteeism
Top Half	61	Top Half	66
Bottom Half	39	Bottom Half	34
	% Below Median on Shrinkage		% Below Median on Shrinkage
Top Half	63	Top Half	69
Bottom Half	37	Bottom Half	31
	% Above Median on Composite Performance (Total)		% Above Median on Composite Performance (Total)
Top Half	68	Top Half	73
Bottom Half	32	Bottom Half	27
	% Above Median on Composite Performance (Direct Outcomes)		% Above Median on Composite Performance (Direct Outcomes)
Top Half	67	Top Half	71
Bottom Half	33	Bottom Half	29

To illustrate this further, Table 7 was prepared to show the probability of above-average performance for various levels of employee engagement. Business units at the highest levels of employee engagement across all business units in Gallup's database have an 83% chance of having high (above average) composite performance. This compares to a 15% chance for those with the lowest levels of employee engagement. So it is possible to achieve high performance without high employee engagement, but the odds are substantially lower (in fact, more than five times lower).

**Table 7 — Percentage of Business Units Above the Median of all Business Units on Composite Performance (Customer, Profit, Productivity, Turnover, Safety, Absenteeism, and Shrinkage) for Different Employee Engagement Percentiles**

<b>Employee Engagement Percentile</b>	<b>Percent Above Median of All Business Units</b>
<b>99th</b>	83%
95th	76%
90th	71%
80th	65%
70th	59%
60th	55%
50th	50%
40th	45%
30th	41%
20th	35%
10th	27%
5th	21%
<b>1st</b>	15%

Other forms of expressing the practical meaning behind the effects from this study include utility analysis methods (Schmidt & Rauschenberger, 1986). Formulas have been derived for estimating the dollar-value increases in output as a result of improved employee selection. These formulas can be used in estimating the difference in performance outcomes at different levels in the distribution of Q<sup>12</sup> scores. Previous studies (Harter, Schmidt, & Hayes, 2002; Harter & Schmidt, 2000) provided utility analysis examples, comparing differences in outcomes between the top and bottom quartiles on the Q<sup>12</sup>. For companies included in this meta-analysis, it is typical to see differences between top and bottom engagement quartiles of 2-4 points on customer loyalty, 1-4 points on profitability, hundreds of thousands of dollars on productivity figures per month, and 4-10 points in turnover for low-turnover companies and 15-50 points for high-turnover companies.

Gallup researchers recently conducted utility analysis across multiple organizations with similar outcome metric types (an update of analyses presented in Harter, Schmidt, & Hayes, 2002, p. 275, Table 6). Comparing top- to bottom-quartile engagement business units resulted in median percentage differences of:

- 31% in turnover for high-turnover companies (those with 60% or higher annualized turnover)
- 51% in turnover for low-turnover companies (those with 40% or lower annualized turnover)
- 12% in customer loyalty/engagement
- 62% in safety incidents
- 51% in shrinkage
- 18% in productivity
- 12% in profitability

Gallup studies conducted at the individual level (rather than the business-unit level) indicate engaged employees have 27% less absenteeism in comparison to disengaged employees.

The above differences and their utility in dollar terms should be calculated for each company, given the company's unique metrics, situation, and distribution of outcomes across business units.

One can see that the above relationships are nontrivial if the business has many business units. The point of the utility analysis, consistent with the literature that has taken a serious look at utility, is that the relationship between employee engagement and business outcomes, even conservatively expressed, is meaningful from a practical perspective.

# Q<sup>12</sup>® Meta-Analysis

## *Discussion*

### **Discussion**

---

Findings reported in this updated meta-analysis provide cross-validation to prior meta-analyses conducted on the Q<sup>12</sup> instrument. The present study expands the size of the meta-analytic database by 10,159 business or work units (an increase of 74%), as well as the number of countries, industries, operating unit types, and outcomes studied. The relationship between engagement and performance at the business-unit level is substantial and highly generalizable across companies. Differences in correlations across companies can be attributed to study artifacts. These findings are important, because they mean generalizable tools can be developed and used across very different organizations, with a high level of confidence that they elicit important performance-related information. The data from the present study provide further substantiation to the theory that doing what is best for employees does not have to contradict what is best for the business.

While the present meta-analysis further substantiates the relationship between employee engagement and performance, additional studies are ongoing that look at issues of causality in more detail. As indicated earlier, Gallup's growing database provides opportunity to design studies that explore causality from many angles. While numerous qualitative analyses are conducted to understand what managers do that effects change in engagement and performance, quantitative studies are also being developed and completed. Such studies, ideally, include elements of meta-analysis (multiple studies), longitudinal analysis (to take into account the nature of time in causality), and path analysis (to take into account the influence of multiple variables). One such study was recently completed, across 2,178 business-units in 10 companies (Harter et al., 2005), testing competing causal models and finding strong evidence for engagement as a leading indicator (as opposed to a trailing result) of business outcomes. However, it is likely that there is some reciprocal relationship between business outcomes and employee engagement. We would expect that in most healthy business units, a reciprocal relationship would exist, in which engagement fuels better business and better business fuels ownership and engagement (involvement and enthusiasm).

The most convincing causal evidence comes, not from one study, but from a body of research and a multitude of types of evidence, including qualitative analysis of high-performing business units, path analysis, predictive studies, and studies of change over time. Such individual studies are a part of Gallup's past and ongoing workplace management research practice.

It is also worth noting that, as Gallup consultants have educated managers and partnered with companies on change initiatives, companies have experienced (between the first and second year), on average, one-half standard deviation growth on employee engagement, and often a full standard deviation growth and more after three or more years. A very important element in the utility of any applied instrument and improvement process is the extent to which the variable under study can be changed. Our current evidence is that employee engagement is changeable, and varies widely by business unit or workgroup.

In addition, work has been done showing that, at the individual level, employee satisfaction is at least somewhat trait related (Arvey, Bouchard, Segal, & Abraham, 1989; Bouchard, 1997). In the present analysis, for business units, we have averaged the independent variable across individuals, which makes our measure more indicative of business-unit performance-related culture than of individual employee satisfaction. In averaging across individuals, we average out trait-related variations, producing a score that reflects the culture of the business unit.

Future research should focus on how employee engagement interacts with other variables to influence outcomes of interest (such as employee talents, customer engagement, length of service, diversity of teams, training, incentive pay, and other change initiatives).

The authors conclude from this study, as with prior Gallup studies, that employee engagement, as measured by the Q<sup>12</sup> items, relates to meaningful business outcomes, and that these relationships can be generalized across companies. The relationships observed are in the directions hypothesized and make psychological sense. Inferences of causality will depend on various pieces of evidence (outlined above) that are collected on an ongoing basis by scientists. In addition, future research published in academic journals may help to shed additional light on the question of causality. Clearly, there are differences across business units in the way employees perceive their work environments, and these differences relate to differences in performance. Such differences represent substantial utility to businesses and other organizations.

# Q<sup>12</sup>® Meta-Analysis

## References

### References

---

- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, 97(1), 129-133.
- Arvey, R. D., Bouchard, T. J., Jr., Segal, N. L., & Abraham, L. M. (1989). Job satisfaction: Environmental and genetic components. *Journal of Applied Psychology*, 74(2), 187-192.
- Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin*, 99(3), 388-399.
- Batt, R. (2002). Managing customer services: Human resource practices, quit rates, and sales growth. *Academy of Management Journal*, 45(3), 587-597.
- Bouchard, T. J., Jr. (1997). Genetic influence on mental abilities, personality, vocational interests, and work attitudes. In C. L. Cooper & I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology*, 12, 373-395. New York: John Wiley.
- Carver, R. P. (1975). The Coleman Report: Using inappropriately designed achievement tests. *American Educational Research Journal*, 12(1), 77-86.
- Denison, D. R. (1990). *Corporate culture and organizational effectiveness*. New York: John Wiley.
- Fleming, J. H. (Ed.). Linkage analysis [Special issue]. (2000, Winter/Spring). *The Gallup Research Journal*, 3(1), 3-138.
- Fleming, J. H., Coffman, C., & Harter, J. K. (2005, July-August). Manage your Human Sigma. *Harvard Business Review*, 83(7), 106-114.
- Gallup, G. H. (1976, Winter). Human needs and satisfactions: A global survey. *Public Opinion Quarterly*, 40(4), 459-467.
- Gallup, G. H., & Hill, E. (1959). *The secrets of long life*. New York: Bernard Geis.
- The Gallup Organization (1992-1999). *Gallup Workplace Audit* (Copyright Registration Certificate TX-5 080 066). Washington, DC: U.S. Copyright Office.

- Grissom, R. J. (1994). Probability of the superior outcome of one treatment over another. *Journal of Applied Psychology*, 79(2), 314-316.
- Harter, J. K., Asplund, J. W., & Fleming, J. H. (2004, August). *HumanSigma: A meta-analysis of the relationship between employee engagement, customer engagement and financial performance*. Omaha, NE: The Gallup Organization.
- Harter, J. K., & Creglow, A. (1997). *A meta-analysis and utility analysis of the relationship between core GWA employee perceptions and business outcomes*. Lincoln, NE: The Gallup Organization.
- Harter, J. K., & Creglow, A. (1998, July). *A meta-analysis and utility analysis of the relationship between core GWA employee perceptions and business outcomes*. Lincoln, NE: The Gallup Organization.
- Harter, J. K., Hayes, T. L., & Schmidt, F. L. (2004, January). *Meta-analytic predictive validity of Gallup Selection Research Instruments (SRI)*. Omaha, NE: The Gallup Organization.
- Harter, J. K., & Schmidt, F. L. (2000, March). *Validation of a performance-related and actionable management tool: A meta-analysis and utility analysis*. Princeton, NJ: The Gallup Organization.
- Harter, J. K., & Schmidt, F. L. (2002, March). *Employee engagement, satisfaction, and business-unit-level outcomes: A meta-analysis*. Lincoln, NE: The Gallup Organization.
- Harter, J. K., & Schmidt, F. L. (2006). Connecting employee satisfaction to business unit performance. In A. I. Kraut (Ed.), *Getting action from organizational surveys: New concepts, technologies, and applications* (pp. 33-52). San Francisco: Jossey-Bass.
- Harter, J. K., Schmidt, F. L., Asplund, J. W., & Killham, E. A. (2005). *Employee engagement and performance: A meta-analytic study of causal direction*. Omaha, NE: The Gallup Organization.
- Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology*, 87(2), 268-279.
- Harter, J. K., Schmidt, F. L., & Killham, E. A. (2003, July). *Employee engagement, satisfaction, and business-unit-level outcomes: A meta-analysis*. Omaha, NE: The Gallup Organization.
- Hunter, J. E., & Schmidt, F. L. (1983). Quantifying the effects of psychological interventions on employee job performance and work-force productivity. *American Psychologist*, 38(4), 473-478.

- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage.
- Hunter, J. E., Schmidt, F. L., & Le, H. A. (in press). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*.
- Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity and corporate financial performance. *Academy of Management Journal*, 38(3), 635-672.
- Iaffaldano, M. T., & Muchinsky, P. M. (1985). Job satisfaction and job performance: A meta-analysis. *Psychological Bulletin*, 97(2), 251-273.
- Johnson, J. W. (1996). Linking employee perceptions of service climate to customer satisfaction. *Personnel Psychology*, 49, 831-851.
- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127(3), 376-407.
- Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48(12), 1181-1209.
- Mayer, R. C., & Schoorman, F. D. (1992). Predicting participation and production outcomes through a two-dimensional model of organizational commitment. *Academy of Management Journal*, 35(3), 671-684.
- Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika*, 8, 161-168.
- National Technical Information Services. (1987). *Standard Industrial Classification manual*. Washington, DC: Executive Office of the President, Office of Management and Budget.
- Ostroff, C. (1992). The relationship between satisfaction, attitudes, and performance: An organizational level analysis. *Journal of Applied Psychology*, 77(6), 963-974.
- Reynierse, J. H., & Harker, J. B. (1992). Employee and customer perceptions of service in banks: Teller and customer service representative ratings. *Human Resource Planning*, 15(4), 31-46.



- Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*, 166-169.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist, 47*(10), 1173-1181.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529-540.
- Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*(2), 199-223.
- Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology, 64*(6), 609-626.
- Schmidt, F. L., Hunter, J. E., Pearlman, K., & Rothstein-Hirsh, H. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology, 38*, 697-798.
- Schmidt, F. L., & Le, H. A. (2004). Software for the Hunter-Schmidt meta-analysis methods. Iowa City, IA: Tippie College of Business, University of Iowa.
- Schmidt, F. L., & Rader, M. (1999). Exploring the boundary conditions for interview validity: Meta-analytic validity findings for a new interview type. *Personnel Psychology, 52*, 445-464.
- Schmidt, F. L., & Rauschenberger, J. (1986, April). *Utility analysis for practitioners*. Paper presented at the First Annual Conference of The Society for Industrial and Organizational Psychology, Chicago, IL.
- Schmit, M. J., & Allscheid, S. P. (1995). Employee attitudes and customer satisfaction: Making theoretical and empirical connections. *Personnel Psychology, 48*, 521-536.
- Schneider, B. (1991). Service quality and profits: Can you have your cake and eat it too? *Human Resource Planning, 14*(2), 151-157.
- Schneider, B., Ashworth, S. D., Higgs, A. C., & Carr, L. (1996). Design, validity, and use of strategically focused employee attitude surveys. *Personnel Psychology, 49*(3), 695-705.
- Schneider, B., & Bowen, D. E. (1993). The service organization: Human resources management is crucial. *Organizational Dynamics, 21*, 39-52.

- Schneider, B., Parkington, J. J., & Buxton, V. M. (1980). Employee and customer perceptions of service in banks. *Administrative Science Quarterly*, 25, 252-267.
- Sechrest, L., & Yeaton, W. H. (1982). Magnitudes of experimental effects in social science research. *Evaluation Review*, 6(5), 579-600.
- Ulrich, D., Halbhook, R., Meder, D., Stuchlik, M., & Thorpe, S. (1991). Employee and customer attachment: Synergies for competitive advantage. *Human Resource Planning*, 14(2), 89-103.
- Wiley, J. W. (1991). Customer satisfaction: A supportive work environment and its financial cost. *Human Resource Planning*, 14(2), 117-127.
- Zohar, D. (1980). Safety climate in industrial organizations: Theoretical and applied implications. *Journal of Applied Psychology*, 65(1), 96-102.
- Zohar, D. (2000). A group-level model of safety climate: Testing the effect of group climate on microaccidents in manufacturing jobs. *Journal of Applied Psychology*, 85(4), 587-596.

# Appendix A

## Reliabilities of Business-Unit Outcomes

---

(based on Schmidt & Hunter, 1996, Scenario 23, p. 219)

Customer		Profitability		Productivity		Turnover		Safety	
Reliab.	Freq.	Reliab.	Freq.	Reliab.	Freq.	Reliab.	Freq.	Reliab.	Freq.
0.89	1	1	3	1	4	1	1	0.63	1
0.87	1	0.99	2	0.99	2	0.62	1	0.82	1
0.84	1	0.93	1	0.92	2	0.6	1	0.66	1
0.75	1	0.91	1	0.9	1	0.39	1	0.84	1
0.58	1	0.9	1	0.62	1	0.27	1		
0.52	1	0.89	2	0.57	1	0.24	1		
0.46	1	0.79	1			0.25	1		
0.33	1	0.57	1			0.63	1		
0.51	1	0.56	1						
0.53	1	0.94	1						

# Appendix B

## Test-Retest Reliabilities of Employee Engagement

---

(based on Schmidt & Hunter, 1996, Scenario 23, p. 219)

Engagement	
Reliability	Frequency
0.92	1
0.86	1
0.83	1
0.80	1
0.79	1
0.78	1
0.77	1
0.66	1
0.82	1
0.97	1
0.65	1
0.45	1
0.75	1