

THE RELATIONSHIP BETWEEN CUSTOMER ENGAGEMENT AND ORGANIZATIONAL OUTCOMES IN THE BUSINESS-TO-BUSINESS CONTEXT

2013 B2B CUSTOMER ENGAGEMENT META-ANALYSIS

Daniela Yu, Ph.D., Gallup

James K. Harter, Ph.D., Gallup

John Fleming, Ph.D., Gallup

JANUARY 2014

The authors thank Jim Asplund, Ed O'Boyle, Julie Lamski, Emily Meyer, Stephanie Plowman, Kirti Kanitkar, Mike Lemberger, Ken Royal, Nancy Marinelli, Aaron Rice, Bill Reid, Lerone Carson, Shay Klevay, Dan Witters, Bill Diggins, Bill Petti, Kyle McKenzie, Marc Thompson, Kevin McConville, Herb Jackson, and Jack Ludwig for contributing important research studies, client data, and database information.

The authors also thank Frank L. Schmidt, Jim Asplund, Sangeeta Agrawal, and Anthony Blue for providing comments on the analysis.

Copyright Standards

This document contains proprietary research, copyrighted materials, and literary property of Gallup, Inc. It is for the guidance of your organization only and is not to be copied, quoted, published, or divulged to others outside of your organization. Gallup®, Gallup Panel™, The Gallup Path®, CE¹¹®, and CE³™ are trademarks of Gallup, Inc. All other trademarks are the property of their respective owners.

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

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

THE RELATIONSHIP BETWEEN CUSTOMER ENGAGEMENT AND ORGANIZATIONAL OUTCOMES IN THE BUSINESS-TO-BUSINESS CONTEXT

2013 B2B CUSTOMER ENGAGEMENT META-ANALYSIS

Daniela Yu, Ph.D., Gallup

James K. Harter, Ph.D., Gallup

John Fleming, Ph.D., Gallup

JANUARY 2014

Table of Contents

Executive Summary	1
Objectives	1
Methods	1
Results	1
Conclusion	2
Development of the Customer Engagement Framework	2
Customer Satisfaction Is Not Enough	2
Advocacy Metrics Are Limited	2
The Truth About Customer Loyalty	3
Measuring What Matters	3
Developing the Best Metrics	4
Description of the CE¹¹ Framework	4
The CE ¹¹ Items	4
Conceptual Hierarchy of the Framework	5

Scientific Evolution and Advancement of Customer Engagement Framework From CE¹¹[®] to CE³[™] . . .	6
Description of the CE ³ Items.	6
Customer Engagement Groups	6
Meta-Analysis, Hypotheses, Methods, and Results	7
Meta-Analysis.	7
Hypotheses and Study Characteristics	7
Meta-Analytic Methods Used	10
Results	12
Utility Analysis: Practicality of the Effects	16
Utility Analysis.	16
Discussion	16
References.	18
Appendix A. Reliabilities of Business/Work-Unit Outcomes	19
Appendix B. Test-Retest Reliabilities of B2B Customer Engagement	19
Appendix C. Between-Company Artifact Distribution of Range-Restriction/Variation Estimates.	20
Appendix D. Within-Company Artifact Distribution of Range-Restriction/Variation Estimates	21

EXECUTIVE SUMMARY

OBJECTIVES

Compared with assessing business-to-consumer (B2C) customer relationships, measuring and managing business-to-business (B2B) customer relationships poses some unique challenges. These include: (1) purchasing cycles that are always long and sometimes inconsistent; (2) the presence of multiple touch points at different levels of the organizational hierarchy in both companies; and (3) the possibility that business customers could be more rational and focus less on emotional connections with their suppliers than consumers do. Due to these unique challenges, this study seeks to determine whether Gallup's customer engagement metric developed in B2C context is a valid and appropriate metric for use in the B2B context. In particular, this study explores whether and to what extent B2B customer engagement relates to key organizational outcomes.

The purpose of this study is thus to examine:

- the true relationship between customer engagement and performance in the B2B context
- the consistency and generalizability of the relationship between customer engagement and performance across organizations in various industries
- the practical meaning of these findings for executives and managers

METHODS

This is the first meta-analysis Gallup has conducted to study the relationship between customer engagement and business performance outcomes in the B2B context. We compiled 23 research studies including three U.S. Gallup Panel studies and nine B2B clients, across six industries and two countries. Within each study, we statistically calculated the business-unit-level relationship between customer engagement and the performance metrics supplied by the various companies. In total, we were able to include 108,989 respondents comprising 19,093 business units in the analysis. We studied six broad classes of outcomes: profitability, revenue/sales, share of wallet,

customer attrition, days sales outstanding (DSO), and brand preference.

Individual studies often contain small sample sizes and idiosyncrasies that distort the interpretation of results. Meta-analysis is a statistical technique that is useful in combining results of studies with seemingly disparate findings, correcting for sampling, measurement error, and other study artifacts to understand the true relationship with greater precision. We applied Hunter-Schmidt meta-analysis methods to the 23 studies to estimate the true relationship between customer engagement and each performance measure and to test for generalizability. After conducting the meta-analysis, we examined the practical effects of the observed relationships by conducting a utility analysis.

RESULTS

Our results demonstrate that in the B2B context, customer engagement is related to each of the six performance outcomes studied. Results indicate high generalizability, which means the correlations are consistent across different organizations. The true score correlations between customer engagement and performance outcomes are 0.24 for profitability, 0.58 for revenue/sales, -0.18 for days sales outstanding, 0.74 for share of wallet, -0.78 for customer attrition, and 0.90 for brand preference. Share of wallet, customer attrition, and brand preference in this study are self-reported outcomes collected during the same survey setting as customer engagement data collection. Thus, the correlations are somewhat higher than correlations between customer engagement measures and variables external to the survey setting. Median performance differences between top-quartile and bottom-quartile units are 34% in profitability, 50% in revenue/sales, 55% in share of wallet, 33% in brand preference, -63% in customer attrition, and -32% in days sales outstanding.¹

¹ For the quantitative analysis in this study, CE^{11®} mean score has been used to calculate the performance gaps between top- and bottom-quartile business units. As Gallup has advanced its science, we recommend that clients use CE^{3™} moving forward. CE³ measures and manages the same construct of customer engagement with greater efficiency and simplicity. The Pearson correlation between mean scores of CE^{11®} and CE³ is 0.95 in B2B context. This almost perfect linear correlation between CE^{11®} and CE³ ensures the results of the quantitative analysis using CE³ will be very close to the results presented here using CE^{11®}.

CONCLUSION

The relationship between B2B customer engagement and performance at the business-unit level is substantial and highly generalizable across organizations. B2B customer engagement is strongly related to each of six different performance outcomes. This means that practitioners can apply Gallup's customer engagement measure in a variety of B2B situations with confidence that the measure captures important performance-related information.

DEVELOPMENT OF THE CUSTOMER ENGAGEMENT FRAMEWORK

While most companies gather data about their customers, few ask the right questions to achieve meaningful performance objectives. Consequently, leaders waste valuable time and resources by pursuing and analyzing the wrong data, which can lead them to make ineffective, costly, or even damaging business decisions. Gallup has studied human behavior and performance for decades, and our extensive customer research provides us with a unique perspective on what works and doesn't work when it comes to surveying customers and improving business performance. Gallup has completed in-depth studies on the most prevalent trends in customer research, including satisfaction, advocacy, and loyalty — and we have found that these models ultimately lead companies to fall short of their goals.

CUSTOMER SATISFACTION IS NOT ENOUGH

Businesses have spent years trying to identify and measure the objective factors that they assume will produce highly satisfied customers. To that end, these efforts have tended to focus on product and price. While these rational factors may indeed satisfy customers to a certain extent, Gallup's research has found that measuring customer satisfaction alone does not generally relate to business outcomes. In studying customers who were extremely satisfied with a product or service (those who provided a rating of "5" on a five-point scale), Gallup found that there were distinct behavioral differences between those whose high satisfaction rating was accompanied by a strong emotional connection (emotionally satisfied customers) and those whose satisfaction ratings weren't (rationally satisfied

customers). Emotionally satisfied customers delivered substantially enhanced value by buying more products, spending more for their purchases, returning more often, or staying longer with the business. Conversely, rationally satisfied customers behaved no differently than customers who were dissatisfied (those who provided a rating of "1" to "4" on a five-point scale) (Fleming & Asplund, 2007).

Finally, satisfaction scores tend to reflect only past experiences and overlook crucial emotional components that drive future purchases. Thus, satisfaction measures tend to show a snapshot of a customer's rational assessment of a brand, service, or product at a particular moment in time, which in itself is not very useful, because it fails to capture intentions or feelings that might predict future behavior.

ADVOCACY METRICS ARE LIMITED

One metric that has recently risen to prominence is a simple, single-item performance measure called the Net Promoter Score (NPS) (Reichheld, 2003). The creator of the NPS construct claims that the only number a company needs to know and track is the percentage of existing customers who would serve as strong advocates for the company (by recommending it to others) minus the company's detractors (customers who would not advocate for the company at all). This metric probably seems like a tempting alternative for business leaders who feel bogged down by complex customer feedback systems — but how simple is too simple?

When it comes to the use of effective metrics, single item measures are inherently less reliable compared with composite or multi-item metrics. Essentially, a single item might not always measure the same things in the same ways from one time to another.

To learn more about the effectiveness of the NPS, Gallup conducted multiple studies looking at three types of customers: **nonadvocates**, who are less than extremely likely to recommend the company to others; **rational advocates**, who, while extremely likely to recommend the company to others, lack a strong emotional bond with the company; and **emotional advocates**, who are also extremely likely to recommend the company to others, but who have forged a strong emotional attachment to the company. Again, Gallup found that the intensity of the customer's emotional

connection to the company matters most. Emotional advocates deliver significantly enhanced business outcomes compared with their rational counterparts. Surprisingly, rational advocates do not differ from nonadvocates on these key measures. To drive advocacy — and reap the financial benefits that come from building strong emotional connections with customers — businesses must first understand and manage those emotional connections (Fleming, 2006).

THE TRUTH ABOUT CUSTOMER LOYALTY

Loyalty has become a common watchword, particularly in bleak economic times, as companies struggle to hang on to the customers they've fought to attract and retain. Most businesses do not understand what makes a customer loyal, yet many invest enormous amounts of money into creating and marketing loyalty programs designed to bring in new customers and keep current ones. Across all industries, companies spend billions each year on loyalty programs. But while businesses pay top dollar to lure customers to join these programs, Gallup's research shows that the programs alone don't change the purchasing behaviors of a large percentage of those whom they enroll.

Gallup determined that loyalty and rewards programs are most successful when they create participants who say they are "much more likely" to shop or use the given brand because of their membership with the brand's loyalty or rewards program and who also say they are "much less likely" or "a little less likely" to use other brands because of their membership with the program. Gallup found that relatively small percentages of shoppers who enroll in loyalty or rewards programs with their preferred brand actually reach this level of loyalty. And even when they do, it doesn't guarantee that they will spend more, unless the company works to build an emotional connection between the customer and the company's program, brand, or product (Ott, 2011).

MEASURING WHAT MATTERS

Companies that seek to connect with their customers solely on a rational level — and to measure their customers' experiences through rational means, such as satisfaction

metrics, loyalty programs, or advocacy scores — are missing an essential piece of the puzzle. Human nature is much more complicated than that, and leaders who ignore its complexities do so at their own peril. By simply maintaining a transactional relationship with their clientele, these companies run the risk that their customers will defect en masse when another business offers a product or service that is less expensive or more convenient. By contrast, Gallup measures not just satisfaction, or loyalty, or advocacy. Instead, Gallup takes a holistic approach that combines elements of all three to measure customer engagement — a customer's emotional or psychological attachment to a brand, product, or company. Gallup's metric uses actionable question items with proven links to customer behavior, representing an important evolution in the way businesses relate to their customers.

In developing this unique approach to customer engagement, Gallup turned to the emerging science of behavioral economics, which theorizes that the vast majority of customer loyalty and buying decisions hinge on emotional, rather than rational, factors. While the prevailing classical economics mindset puts reason at the center of people's fiscal actions, behavioral economists believe that rational considerations actually account for less than one-third of human decisions and behaviors. Metaphorically speaking, this means that the majority of a customer's buying decisions are made more from the heart than from the head. This revolutionary finding has the power to turn the business world — which largely abides by the tenets of classical economics — upside down. Gallup believes that for customers, feelings are facts and perceptions are reality, and any metric that does not account for this aspect of human nature is fundamentally flawed. This breakthrough led Gallup researchers to uncover what the purely rational customer measurements are missing: They do not take into account the deeper psychological reasons that draw people to a product or service or keep them coming back to a company. Gallup's research has clearly shown that the secret to winning and keeping customers is to measure and manage their emotional engagement with a brand.

DEVELOPING THE BEST METRICS

Customer metrics must pass two important tests. First, metrics must be clear and straightforward. Everyone in the organization must understand and relate to the measurements, and they should be easy to communicate to the outside marketplace. Second, metrics must have strong links to key business outcomes. Spending time and effort conducting measurements that are not connected to business performance is common, and wasting resources measuring something that doesn't matter is worse than conducting no measurement at all. The first test is relatively easy to pass. Many metrics used today are so compellingly simple that they seem too good to be true. And smart organizations realize they often *are* too good to be true, because the second test is much more difficult to pass. Studies, including Gallup's research, show that many popular metrics — like customer satisfaction, customer loyalty, and advocacy (including Net Promoter) — do not consistently demonstrate strong links to key business outcomes (Fleming & Asplund, 2007).

With this in mind, Gallup set out to measure customer engagement by testing an exhaustive set of potential questions drawn from previous surveys and academic studies of the psychology of human emotion. Researchers asked customers of different products and services to rate their current or most-used provider on a list of more than 60 candidate measures. Extensive analysis of these measures yielded a final list of eight emotional attachment items that showed strong linkage to attitudinal loyalty and key business performance outcomes. The final set consists of 11 customer engagement items — Gallup's CE¹¹. These items are arranged into two distinct elements that parallel Adam Smith's and Antonio Damasio's dual-process models. One element is arguably more rational (the three "behavioroid" items that measure attitudinal loyalty), and one is more emotional (the eight items that measure emotional attachment). Together, these 11 items assess how engaged a customer is with a particular company.

DESCRIPTION OF THE CE¹¹® FRAMEWORK

In designing the CE¹¹ items, researchers considered the emotional attachment element of human behavior.

THE CE¹¹ ITEMS

CE1. Taking into account all the products and services you receive from them, how satisfied are you with *Company* overall?

CE2. How likely are you to continue to do business with *Company*?

CE3. How likely are you to recommend *Company* to a friend or associate?

CE4. *Company* is a name I can always trust.

CE5. *Company* always delivers on what they promise.

CE6. *Company* always treats me fairly.

CE7. If a problem arises, I can always count on *Company* to reach a fair and satisfactory resolution.

CE8. I feel proud to be a *Company* customer.

CE9. *Company* always treats me with respect.

CE10. *Company* is the perfect company for people like me.

CE11. I can't imagine a world without *Company*.

Respondents rate the CE¹¹ items using six response options. CE1 is measured using a five-point Likert *satisfaction* scale (where "5" means "Extremely satisfied" and "1" means "Not at all satisfied"). CE2 and CE3 use a five-point Likert *likelihood* scale (where "5" means "Extremely likely" and "1" means "Not at all likely"), and CE4 through CE11 use a five-point Likert *agreement* scale (where "5" means "Strongly agree" and "1" means "Strongly disagree").

For all scales, a sixth response option ("Don't know/Does not apply") is unscored.

The combined scale has acceptable internal consistency — Cronbach's alpha for the full 11-item set is 0.94. For the

three attitudinal loyalty items, alpha is 0.79; for the eight emotional attachment items, alpha is 0.92.

CONCEPTUAL HIERARCHY OF THE FRAMEWORK

This section outlines the conceptual framework of Gallup's CE¹¹ customer engagement construct.

The first three items combined measure the rational, loyal attitude a customer has toward a company. The first item is a traditional measure of customer satisfaction. The second item measures a customer's intent to continue to purchase products or services from this company, and the third item measures whether the customer will recommend this company to others.

CE1. Overall satisfaction

CE2. Likelihood to continue

CE3. Likelihood to recommend

The first and foundational dimension of emotional attachment — **confidence** — is the most basic expectation of customers. Gallup research has shown that companies that fail to build confidence have a much tougher time engaging customers than companies in which customers' confidence is solid. One of the main things that people look for in a brand or a company is a promise of consistently high performance. Great companies always keep their promises. Their customers feel secure in dealing with the company, and they don't expect unpleasant surprises. Including the word "always" in these items makes each a difficult standard to live up to, but it represents the gold standard. Companies that *always* keep their promises create trust. Their customers sense that promises made today will be kept in the future. If "sometimes" or "usually" is the best a company can do, that company will move outside its customers' trust circle, and it will be extremely difficult to get back in. In addition, if the company's promise to its customers is not clearly understood by the company's rank-and-file employees, customers may not agree that the company *always* delivers on its promises, thus undermining trust.

CE4. Name I can always trust

CE5. Always delivers on promises

The second dimension of emotional attachment is **integrity**, which is the essential element of equitable treatment. Will a company treat all of its customers fairly, even if doing so costs the firm money in the short term? If something goes awry, can customers count on the company to fix it — and fast? These are the kinds of questions that illustrate customers' expectations regarding fairness.

Gallup research shows that customers who experience a problem typically have significantly lower levels of emotional attachment with the company than those who say they've had no problems. On the other hand, customers who have encountered a problem but are extremely satisfied with the way the company handled (not resolved) the problem often have levels of emotional attachment equal to — and in some cases, exceeding — those who had no problem at all. Handling a problem well can turn a crisis into an opportunity to engage the customer.

CE6. Always treats me fairly

CE7. Always count on fair resolution of any problems

The next dimension, **pride**, denotes a sense of positive association and identification with a company. Pride goes well beyond basic notions of self-presentational, status, or badge qualities of association with a company to deeper levels of shared values between the customer and the company, as well as feelings of being competent and in control. Customers' associations with companies not only convey information about them to others, but those associations also define, shape, and sharpen their own self concepts. Customers need to feel respected and valued in order to be willing to positively associate themselves with a particular brand.

CE8. Feel proud to be a customer

CE9. Always treats me with respect

The final dimension — **passion** — represents the ultimate expression of emotional attachment. Customers who are passionate about a company would consider it a perfect fit for them and irreplaceable. For these customers, the number of potential alternatives they would consider using is zero. They can no more imagine their lives without the company

than they could imagine life without air or water. These customers are relatively rare, but they are *customers for life* and are a significant financial resource and annuity for companies.

In addition, not only are they truly committed “ambassadors” for the brand, but they hold the key to building greater levels of passion throughout a company’s customer population.

CE10. Perfect company for people like me

CE11. Can’t imagine a world without

SCIENTIFIC EVOLUTION AND ADVANCEMENT OF CUSTOMER ENGAGEMENT FRAMEWORK FROM CE¹¹® TO CE³™

The Gallup CE³ represents an evolution and advancement of our science. It measures and manages the same construct of customer engagement as CE¹¹, but with greater efficiency, simplicity, and effectiveness. The new metric does not deny the validity of our past science, but rather uses it as a guide to inform the present.

DESCRIPTION OF THE CE³ ITEMS

CE³ focuses on measuring the emotional attachment of human behavior. After investigating many strategic and statistical parameters, Gallup researchers determined that the final set of items included in the customer engagement instrument would be:

Q1. Company always delivers on what they promise.

Q2. I feel proud to be a Company customer.

Q3. Company is the perfect company for people like me.

Respondents rate the CE³ items using six response options. All three questions use a five-point Likert agreement scale (where 5 means “Strongly agree” and 1 means “Strongly disagree”). For all scales, a sixth response option (“Don’t know/Does not apply”) is unscored.

The combined scale has very high internal consistency — Cronbach’s alpha for the full three-item set is 0.89. These items were found to fulfill the most parameters, making

them the most powerful predictors of CE¹¹ mean and thus, of business outcomes. The Pearson correlation between CE¹¹ mean and CE³ mean is 0.95 in the B2B context, and 0.93 in the B2C context. The almost perfect linear correlation between CE¹¹ and CE³ ensures that these two metrics work in a very similar way, but CE³ is compact and economical.

CUSTOMER ENGAGEMENT GROUPS

As part of the customer engagement instrument, customers with positive, neutral, and negative sentiments are now classified as “Fully Engaged,” “Indifferent,” or “Actively Disengaged.” **Fully Engaged** customers have a strong positive emotional attachment to a company. Some might even say they love that company. They are the company’s most valuable customers; they spend more, visit more often, and stay longer with the company.

Indifferent customers are neutral; they don’t really care about a company one way or the other. They neither love nor hate the brand. As a result, they have no particular allegiance to it and may switch to a different company or brand if opportunity permits and the potential benefits outweigh the costs of switching.

Actively Disengaged customers feel strongly negative about a company. Some might even say they hate that company. They can be actively antagonistic, spread negative word of mouth, and cost the company money. These customers remain because the costs of switching are too high or they feel there are no better options available.

This classification scheme produced differentiated business impacts on organizations across a wide variety of settings (e.g., industries, countries, customer types). To learn more about the research and development of the CE³ metric, please see Gallup’s Customer Engagement Instrument Technical Report (Fleming & Yu, 2014).

In this study, CE¹¹ has been used as the metric for the quantitative analysis. Because of the almost perfect linear correlation between CE¹¹ and CE³, the results based on CE³ will be very similar to those based on CE¹¹. We explored all of the available business metrics provided by client organizations, which conceptually grouped into six classes of business metrics. This meta-analysis includes all available

Gallup studies (whether published or unpublished) and has no risk of publication bias.

This is the first meta-analysis Gallup has conducted of the relationship between customer engagement and business performance metrics in the B2B context. We explored all of the available business metrics provided, and decided to analyze six classes of business metrics that were most consistently provided by our clients. This meta-analysis includes all available Gallup studies (whether published or unpublished) and has no risk of publication bias.

META-ANALYSIS, HYPOTHESES, 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 bias 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, 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, industries, 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 background information and detailed descriptions of the more recent meta-analytic methods: Schmidt (1992); Hunter & Schmidt (1990, 2004); Lipsey

& Wilson (1993); Bangert-Drowns (1986); and Schmidt, Hunter, Pearlman, & Rothstein-Hirsh (1985).

HYPOTHESES AND STUDY CHARACTERISTICS

The hypotheses examined in this meta-analysis are as follows:

- Hypothesis 1: Business-unit-level B2B customer engagement will have positive average correlations with business-unit outcomes of revenue/sales, profitability, share of wallet, and brand preference, and negative correlations with customer attrition and days sales outstanding.
- Hypothesis 2: The correlations between B2B customer engagement and business-unit outcomes will generalize across organizations for all outcomes. That is, these correlations will not vary substantially across organizations. And in particular, there will be few, if any, organizations with zero correlations or correlations in the opposite direction from Hypothesis 1.

Gallup's B2B customer engagement database includes 815 studies conducted as proprietary research for 185 independent organizations. In each study contained in the database, customer engagement data for all respondents are aggregated at the business-unit level. Not all of the studies in the database, however, have associated business performance metrics provided by the respective clients.

In the present study, customer engagement data and associated business performance metrics were obtained for 23 of the most recent studies in the database and aggregated at the business-unit level. As a result, the unit of analysis in this study is the business unit. Customer engagement data from the 23 studies were correlated with the following classes of aggregate business-unit performance measures:

- profitability
- brand preference
- revenue/sales
- customer attrition

- share of wallet
- days sales outstanding

Gallup's database contains three types of studies. In the first type, **single-contact design**, a client's B2B accounts are each assigned just one contact to complete the survey. As a result, that individual completing the survey represents all of the members of the account team. Usually that person is the account manager or account representative who rates the relationship with the client on behalf of the entire team. Examples of single-contact design studies might include: a hotel property owner rating his or her engagement with a large hotel chain; or a physician in a private practice rating his or her engagement with a medical device provider.

In the second type of study, **multiple-contact design**, a client's B2B accounts are each assigned multiple contacts to complete the survey, and an attempt is made to interview as many of these contacts as possible. The contacts generally represent four different classes of team members, which together comprise a buying center: decision makers, influencers, buyers, and end users. Examples of multiple-contact design studies might include: all of the members of a company's project team and executive committee rating their engagement with a consulting company; or a company's leadership and middle management teams evaluating their engagement with an outsourced provider.

Occasionally, however, due to difficulties in reaching all of the identified contacts, a multiple-contact design study might become a single-contact study because only one individual actually provides a rating. This is the third type of study: **multiple-contact design in theory, but single-contact in reality**. As we have noted, no matter which type of study design, we aggregate the CE³ scores in the analysis (and in our database) at the business-unit level. It is conceivable that these variants of B2B study design could produce different relationships between CE³ scores and business performance metrics. Consequently, we performed the meta-analysis on single- and multiple-contact design studies separately to determine whether the relationship between metrics differed. Although we discuss the results

for both types below, the data suggest that there are no meaningful differences between the two types of study design — so in the main analysis, all studies were analyzed together.

This paper does not directly speak to issues of causality, which are best addressed using meta-analytic longitudinal data, consideration of multiple variables, and path analysis. Since there are not enough meta-analytic longitudinal data, causality has to be examined in future research. However, it's reasonable to hypothesize that high and/or growing levels of customer engagement result in better business performance, rather than the opposite. When customers are fully engaged (i.e., they are rationally satisfied with and have a strong emotional connection to the company), they are more likely to do more business with the company than customers who are not fully engaged. On the contrary, customers who have done more business with the company are not necessarily more engaged than customers who have conducted less business there.

Studies for the current meta-analysis were selected so that each organization is represented once in each analysis. For several organizations, multiple studies were conducted. To include the best possible information for each organization represented in the study, some basic rules were used. If two concurrent studies were conducted for the same client (where customer engagement and outcome data were collected concurrently — i.e., in the same year), then a composite of effect sizes across the multiple studies was calculated and entered as the value for that organization. If an organization had a concurrent *and* a predictive study (where the customer engagement was collected in Year 1 and outcomes were tracked in Year 2), then the effect sizes from the predictive study were entered. Based on the basic rules above, we calculated three types of correlations. First, Pearson correlations were calculated to estimate the relationship between business-unit-level aggregate measures of customer engagement mean score and each of the six general classes of outcomes. We computed correlations across business/work units in each company. Second, for the same business metrics across years (e.g., Year 1 sales, Year 2 sales) within the same company, we calculated a weighted

average correlation to create a single correlation coefficient across years. Third, for different business metrics that belong to the same outcome category within a client (e.g., sales and net contribution per employee are two different business metrics in the “revenue/sales” category), a composite score correlation R_{xy} was calculated based on equations 1 to 3). R_{xx} is the intercorrelation among business metrics within the same category. Depending on the unique nature of data from each client, the resulting correlation coefficients were entered into a database.

$$\text{Equation 1: } R_{xx} = \frac{1}{n} \sum_{1}^n R_{n,n+1}$$

$$\text{Equation 2: } C_{xx} = \frac{1+(n-1)R_{xx}}{n}$$

$$\text{Equation 3: } R_{xy} = \frac{R_{xx}}{\sqrt{C_{xx}}}$$

The researchers then calculated mean validities, standard deviations of validities, and validity generalization statistics for each of the six business-unit-outcome measures.

- Revenue/sales studies were available for 16 organizations across 9,310 business units. Measures of business-unit revenue/sales consisted of one of the following: financials (e.g., revenue/sales dollars per branch or store, cost effectiveness) or quantity produced (e.g., production volume, number of accounts opened in banks, number of medical devices purchased). The majority of variables included as “revenue/sales” were financial measures of revenue/sales or growth in revenue/sales.
- Profitability studies were available for four organizations across 916 business units. Measures of business-unit profitability consisted of percentage profit of revenue (sales) — i.e., measures of margin rather than financial dollars.
- Share-of-wallet studies were available for three organizations across 4,690 business units. Share of wallet is defined as the percentage of a customer’s spending at a given company during a certain period of time — for example, the percentage of nights spent at one hotel brand out of the total number of nights spent at hotels per year, or the percentage of money spent at one grocery store out of the total amount spent at grocery stores per month. Based on this

Table 1 provides a summary of organizations sorted by industry. There is considerable variation in the industry types represented, as organizations from six industries provided studies. These industries are Business Banking, Medical Devices, Industrial Supply, Hospitality, Food Supply, and Professional Services.

Table 1: Summary of Studies by Industry

Industry type	Number of organizations	Number of business units/projects	Number of respondents
Business Banking	1 client, 3 Panel studies	7,640	93,770
Medical Devices	2 clients	4,952	4,952
Industrial Supply	2 clients	5,263	5,263
Hospitality	2 clients	769	3,603
Food Supply	1 client	24	272
Professional Services	1 client	445	1,129

definition, measures of share of wallet consisted of percentages ranging from 0% to 100%.

- Customer attrition (turnover) data were available for three organizations across 4,751 business units. The attrition measure consisted of the annualized percentage of customer turnover for each business unit.
- Brand preference data were included for eight organizations across 4,639 business units. Brand preference is measured by the likelihood of a respondent to choose the current brand as his or her first choice for future business.
- Days sales outstanding (DSO) data were included for two organizations across 615 business units. DSO is a measure of the average number of days that a company takes to collect revenue after a sale has been made. A low DSO number means that it takes a company fewer days to collect its accounts receivable. A high DSO number shows that a company is selling its product to customers on credit and taking longer to collect money.

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 variable, and range variation and restriction in the independent variable (customer engagement mean score) 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 & Schmidt (1990, 2004) include correction for measurement and statistical artifacts such as range restriction and measurement error in the performance variables gathered. The following sections provide definitions for these procedures.

Gallup researchers gathered performance-variable data for multiple time periods to calculate the reliabilities of the performance measures. Because these multiple measures were not available for each study, the researchers used

artifact distribution-based meta-analysis methods (Hunter & Schmidt, 1990, 2004) 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/work-unit outcome measure reliabilities was consistent with Scenario 23 in Schmidt & 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:

$$\text{Equation 4: } R = \frac{R_{12} \times R_{23}}{R_{13}}$$

Where R_{12} is the correlation of the outcome measured at Time 1 with the same outcome measured at Time 2; R_{23} is the correlation of the outcome measured at Time 2 with the outcome measured at Time 3; and R_{13} 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 1 to Time 3 than from Time 1 to Time 2 or from Time 2 to Time 3) from random changes in business-unit results caused by measurement error, sampling error (primarily in customer and quality measures), data collection errors, 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 consistent with 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 profitability, revenue/sales, and customer attrition. But they were not collected for share of wallet, brand preference, and DSO because they were not available. Therefore, the assumed reliability for share of wallet, brand preference, and DSO were 1.00, resulting in downwardly biased true validity estimates (the estimates of validity reported here are lower than reality). Artifact distributions for these three variables will be added to upcoming reports as they become available.

It could be argued that because the independent variable (customer engagement, as measured by the CE¹¹) is used

in practice to predict outcomes, the practitioner must live with the reliability of the instrument he or she is using. However, measurement error in the independent variable, on the one hand, is caused by imperfect measurement through item coverage and transient error, and on the other by limited sample size per business unit (due to financial budgets and practical concerns). Correcting for independent measurement error provides some guidance in how strongly the true score variables relate to one another. In other words, because it is not feasible to collect data on all customers and measure their engagement perfectly, the true score estimate provides guidance in the strength of the relationship. Appendix B presents the distributions of reliabilities for the CE¹¹ mean score. These values were computed in the same manner as were those for the performance 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). Similar to what has been argued in the employee engagement arena (Harter, Schmidt, Agrawal, & Plowman, 2013), one could argue that there is no explicit range restriction because we are studying results as they exist in the business units or projects and they are not selected based on scores on the predictor (customer engagement mean score). However, in studying companies, we have observed that there is variation across companies in standard deviations of indexes. Therefore, the standard deviation of the population of business units across organizations studied will be greater than the standard deviation within a 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, 2006). One hypothesis for why this variation occurs is that companies vary in how they encourage customer engagement initiatives and in their capabilities to actually improve customer engagement.

Since the initial development of Gallup's customer engagement metric and its application to B2B customer relationships, Gallup has collected descriptive data on more than 984,264 respondents in 12,110 business units from 275 organizations. This accumulation of data indicates that the standard deviation within a given company is approximately 90% of the standard deviation in the population of all organizations.

In addition, the ratio of the standard deviation for a given organization relative to the population value varies from organization to organization. Therefore, if one goal is to estimate the effect size in the population of all business units (arguably an important issue), then correction should be made based on such available data. In the observed data, correlations are attenuated for organizations with less variability across business/work units than the population average, and vice versa. As such, variability in standard deviations across organizations will create variability in observed correlations and is therefore an artifact that can be corrected for in interpreting the generalizability of validities. Appendix C presents artifact distribution for range-restriction/variation corrections used for meta-analysis. We have included all B2B customer engagement client organizations from 2008 to 2012. They resemble those reported in the earlier study, but with a larger number of entries. The following excerpt from Hunter & Schmidt (2004) provides an overview of meta-analysis using artifact distributions:

In any given meta-analysis, there may be several artifacts for which 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 for 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 measurement error and range variation (Hunter & Schmidt, 1990, 2004).*

In this study, statistics were calculated and reported at each level of analysis, starting with the observed correlations, and then corrected for sampling error, measurement error, and range variation. Both within-organization range-variation corrections (to correct validity generalization estimates) and between-organization range-restriction corrections (to correct for differences in variation across organizations) were made. Between-organization range-restriction corrections are relevant in understanding how engagement relates to performance across business/work units in all organizations. As alluded to, we have applied the indirect range-restriction correction procedure to this meta-analysis (Hunter et al., 2006).

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 based on 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. (2006) technique:

$$\text{Equation 5: } S_e^2 = (1 - \bar{r}^2)^2 / (\bar{N} - 1)$$

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 of 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. Generalizability is generally assumed if a high percentage (such as 75%) of the variance in validities across

studies is due to sampling error and other artifacts, or if the 90% credibility value (10th percentile of the distribution of true validities) is in the hypothesized direction.

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

RESULTS

The focus of analyses for this report is on the relationship between overall customer engagement and a variety of outcomes. Table 2 provides the updated meta-analytic and validity generalization statistics for the relationship between customer engagement and performance for each of the six outcomes studied. Two forms of true validity estimation follow mean observed correlations and standard deviations. The first corrects for range restriction across the population of business/work units, independent-variable measurement error, and dependent-variable measurement error. Estimates that include the range-restriction correction apply to interpretations of effects in business/work units across organizations, as opposed to effects expected within a given organization. The second corrects for range variation within organizations, independent-variable measurement error, and dependent-variable measurement error. This range-variation correction places all organizations on the same basis in terms of variability of customer engagement across business/work units. These results can be viewed as estimating the relationships across business/work units within the average organization, as opposed to effects expected across organizations. Because there is more variation in engagement for business/work units across organizations than there is within the average organization, effect sizes are higher when true validity estimates are calculated for business/work units across organizations.

For instance, observe the estimates relative to the revenue/sales criteria in Table 2. With the “between companies” range-restriction correction (which is relevant to business/work units across organizations), the true validity value of customer engagement is 0.51 with a 90% credibility value (CV) of 0.27. In contrast, with the “within companies”

range-variation correction (which is relevant to business/work units within organizations), the true validity value of customer engagement is 0.20 with a 90% CV of 0.15.

The findings in Table 2 show high generalizability across organizations in the relationship between customer engagement and the performance outcomes measured (profitability, revenue/sales, share of wallet, brand preference, customer attrition, and days sales outstanding). Most of the variability in correlations across organizations was the result of sampling error in individual studies; for each of these six outcomes (except revenue/sales), more than 75% of the variability in correlations across organizations can be attributed to artifacts (sampling error, range

variation, and measurement error). In other words, the true validity is very similar and in the hypothesized direction for each organization studied. For revenue/sales, results indicate high generalizability across the organizations studied, as implied by the 90% credibility value in the hypothesized direction. However, artifacts do not explain all of the variance in correlations of customer engagement and revenue/sales. Regardless, the 90% credibility values indicate substantial evidence of generalizability for all six outcomes studied (Schmidt & Hunter, 1977). What this means is that customer engagement effectively predicts these outcomes in the expected direction across organizations, including those in different industries and in different countries.

Table 2: Meta-Analysis Results of Relationship Between Customer Engagement and Business-Unit-Level Performance²

	Profitability	Revenue/ Sales	Share of wallet	Brand preference	DSO	Attrition	
No. of business units	916	9,310	4,690	5,569	615	4,751	
No. of Rs	4	16	3	10	2	3	
Mean observed R	0.07	0.21	0.31	0.45	-0.06	-0.23	
Observed SD	0.06	0.16	0.01	0.07	0.02	0.02	
Range restriction (between companies)	True Validity	0.21	0.51	0.65	0.79	-0.16	-0.68
	True Validity SD	0.00	0.19	0.00	0.00	0.00	0.00
	% Variance accounted for sampling error	106	6	269	22	948	113
	% Variance attributable to all statistical artifacts	163	62	11,895	627	1,462	3,038
	90% CV	0.21	0.27	0.65	0.79	-0.16	-0.68
	p (true score correlation)	0.24	0.58	0.74	0.90	-0.18	-0.78
	SD of p	0.00	0.22	0.00	0.00	0.00	0.00
Range variation (within companies)	True Validity	0.06	0.20	0.27	0.41	-0.05	-0.29
	True Validity SD	0.00	0.09	0.00	0.00	0.00	0.00
	% Variance accounted for sampling error	106	6	269	22	948	113
	% Variance attributable to all statistical artifacts	169	70	13,474	681	1,548	51,667
	90% CV	0.06	0.15	0.27	0.41	-0.05	-0.29
	p (true score correlation)	0.10	0.33	0.44	0.67	-0.08	-0.48
	SD of p	0.00	0.14	0.00	0.00	0.00	0.00

R=Correlation

SD=Standard Deviation

CV=Credibility Value

² For the quantitative analysis in this study, CE^{11*} mean score has been used to calculate the performance gaps between top- and bottom-quartile business units. As Gallup has advanced its science, we recommend that clients use CE³ moving forward. CE³ measures and manages the same construct of customer engagement with greater efficiency and simplicity. The Pearson correlation between mean scores of CE^{11*} and CE³ is 0.95 in B2B context. This almost perfect linear correlation between CE^{11*} and CE³ ensures the results of the quantitative analysis using CE³ will be very close to the results presented here using CE^{11*}.

As shown in Table 2, the strongest effects were found for brand preference, share of wallet, and customer attrition. These three outcomes are self-reported outcomes from the same survey environment in which the respondent was asked the customer engagement items. Therefore, self-reported outcomes are more likely to produce inflated estimates due to mono-method bias. If respondents report high ratings on customer engagement conditions, they are also more likely to be internally consistent with themselves, thus reporting higher share of wallet, higher likelihood of brand preference, and less likelihood of attrition. The correlation between engagement and revenue/sales is stronger than the correlation between engagement and profitability. Revenue/sales is the direct consequence of customer engagement, while profitability is a downstream result of the intermediary outcome (i.e., revenue/sales), and includes many additional factors such as expenses. Compared with the true correlations between employee engagement and revenue/sales and profitability, customer engagement has a stronger correlation to business outcomes than employee engagement does (Harter et al., 2013). Based on The Gallup Path,³ employee engagement directly affects customer engagement, and as such indirectly affects business performance outcomes such as revenue/sales and profitability — via customer engagement and other mediators such as employee retention and absenteeism.

This explains customer engagement's stronger correlation to these two performance outcomes.

For the revenue/sales outcome, data are sufficient to explore whether differences exist between single- and multiple-respondent studies. The findings show that the correlation to revenue/sales is stronger for single respondents than it is for multiple. In the case of a single respondent, such as a physician, he or she is almost always the sole decision maker on purchasing medical devices if the budget is fixed. If he or she is engaged, he will buy more devices from the current supplier. In the case of multiple respondents, however,

customer engagement is more of an overall reflection of all touch points. While influencers, buyers, and end users are present in the group, these respondents might not all be exactly within the same group of decision makers. Therefore, the number of respondents is a moderator in the correlation between customer engagement and business performance outcomes, with higher correlation in the case of a single respondent.

The statistical correlations substantiate the positive relationship between B2B customer engagement and performance outcomes. The next section will explore the practical utility of the observed relationships.

³ The Gallup Path is a behavioral economic model for organizations. It illustrates the proven, revolutionary strategies used by successful businesses. The Path shows that when employees are engaged, they create engaged customers, which further leads to sustainable growth of business and ultimately real profit increase. For more information, please read *Follow This Path: How the World's Greatest Organizations Drive Growth by Unleashing Human Potential* by Curt Coffman and Gabriel Gonzalez-Molina.

UTILITY ANALYSIS: PRACTICALITY OF THE EFFECTS

UTILITY ANALYSIS

The research literature consulted 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; Harter, Schmidt, & Hayes, 2002). We found the same to be true in this study.

There are many forms of expressing the practical meaning of effect size, such as BESDs (Rosenthal & Rubin, 1982; Grissom, 1994), Schmidt & Rauschenberger's utility analysis (1986), and Gallup's utility analysis of employee engagement (Harter et al., 2002). This study presents practical meaning of effect size in a consistent manner with Gallup's workplace practice (i.e., comparing differences in outcomes between the top and bottom quartiles on the customer engagement mean score).

Comparing top-quartile-engagement business units/projects with those in the bottom quartile resulted in median percentage differences⁴ of:

- 34% in profitability
- 50% in revenue/sales
- 55% in share of wallet
- 33% in brand preference
- -63% in customer attrition
- -32% in days sales outstanding

These differences and their utility in financial terms should be calculated for each organization, given the organization's unique metrics, situation, and distribution of outcomes across business units. The median estimates represent the midpoint in the distribution of utility analyses conducted across many studies (49 for revenue/sales, 11 for profitability, nine for share of wallet, 11 for brand preference, three for customer attrition, and two for days sales outstanding), depending on the outcome and availability of organizational data with similar outcome types.

One can see that the above relationships are nontrivial if the business has many business units or projects. The point of the utility analysis, consistent with literature that has taken a serious look at utility, is that the relationship between B2B customer engagement and organizational outcomes, even conservatively expressed, is meaningful from a practical perspective.

DISCUSSION

The findings from this meta-analysis demonstrate substantial and highly generalizable relationships between customer engagement and six business-unit-level performance indicators across companies in different industries and countries. Differences in correlations across companies can be attributed primarily to study artifacts. These findings are important because they demonstrate that Gallup's customer engagement metric is a generalizable tool that can be used across different organizations with a high level of confidence that it links to important performance-related information. The findings from the present study further substantiate the notion that engaging customers both rationally and emotionally has great practical bottom-line value for organizations.

⁴ For the quantitative analysis in this study, CE^{11*} mean score has been used to calculate the performance gaps between top and bottom quartile business units. As Gallup has advanced its science, we recommend that clients use CE³ moving forward. CE³ measures and manages the same construct of customer engagement with greater efficiency and simplicity. The Pearson correlation between mean scores of CE^{11*} and CE³ is 0.95 in B2B context. This almost perfect linear correlation between CE^{11*} and CE³ ensures the results of the quantitative analysis using CE³ will be very close to the results presented here using CE^{11*}.

Additionally, this is the first meta-analysis that thoroughly investigated whether Gallup's customer engagement metric applies in the B2B context. Contrary to the oft-heard claim that business customers are much more rational in their decision making than are consumers and do not build emotional connections with the companies they do business with, our study shows that emotional connections still matter in B2B customer relationships. Anecdotally, this is revealed in the common explanation offered for preferring one supplier over another — "I just have a greater comfort level with them." In a recent study in Germany,⁵ for example, 77% of B2B decision makers indicated that they disregard emotions in purchasing decisions. Seventy percent claimed to rely only on objective facts when making purchasing decisions. However, a majority of the respondents (54%) also said that they would let a deal fall through (despite a favorable set of objective facts), if they had a "bad feeling" about the deal. Even among German technical and engineering purchasers — arguably a highly rational subgroup of professionals — this suggests that the final choice of suppliers is much more than a rational decision. In other words, even for the most "rational" of purchasers, feelings are facts. The meta-analysis demonstrates that although a B2B relationship is a

relationship between two companies, it is those companies' *people* who fundamentally operationalize that relationship. Hence, any customer measurement approach that fails to take into account the emotional connections that customers form, and focuses myopically on product or price, will not be positioned for greater success going forward.

Because B2B purchases typically have longer cycles than consumer purchases, and because improvements in customer engagement require changing perceptions across multiple touch points in the purchasing organization, it's understandable that improvements in customer engagement take longer in the B2B than the B2C context. Nevertheless, by aggressively implementing change initiatives focused on improving customer engagement, the top 25% of B2B companies in Gallup's database have experienced, on average, one standard deviation growth on CE³ scores after four or more years. An 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 customer engagement is changeable and varies widely by business unit and organization.

⁵ In July/August of 2010, 300 telephone interviews were conducted with persons of the first and second management levels (Owner, General Manager, Directors of Purchasing and Production) of German companies in mechanical engineering, electronics, and automation. These individuals were questioned on their motives in purchasing decisions. Companies in these industries are technically driven and are therefore considered to be highly rational in their decision-making.

REFERENCES

- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, 97(1), 129-133.
- Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin*, 99(3), 388-399.
- B-t-B-Entscheider handeln emotionaler, als sie meinen. (2010). *W&V*, 9, 12.
- Carver, R. P. (1975). The Coleman Report: Using inappropriately designed achievement tests. *American Educational Research Journal*, 12(1), 77-86.
- Coffman, C. W., & Gonzalez-Molina, G. (2002). *Follow This Path: How the World's Greatest Organizations Drive Growth by Unleashing Human Potential*. Gallup Press.
- Fleming, J. H. (2006). A popular idea that's dead wrong. *Gallup Business Journal*. Retrieved from <http://businessjournal.gallup.com/content/25822/PopularIdea-That's-Dead-Wrong.aspx#4>.
- Fleming, J. H., Coffman, C. W., & Harter, J. K. (2005). Manage your Human Sigma. *Harvard Business Review*, 83(7), 106-114.
- Fleming, J. H., & Asplund, J. (2007) *Human Sigma: Managing the Employee-Customer Encounter*. New York: Gallup Press.
- Gallup. (2013). *State of the American Workplace*. Omaha, NE: Gallup, Inc.
- 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., Agrawal, S., & Plowman, S.
- K. (2013). *The Relationship Between Engagement at Work and Organizational Outcomes*. Omaha, NE: Gallup, Inc.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of metaanalysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of metaanalysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage.
- Hunter, J. E., Schmidt, F. L., & Le, H. A. (2006). *Implications of direct and indirect range restriction for metaanalysis methods and findings*. *Journal of Applied Psychology*, 91, 594-612.
- 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.
- Ott, B. (2011). Making loyalty programs work. *Gallup Business Journal*. Retrieved from <http://businessjournal.gallup.com/content/149570/Making-Loyalty-ProgramsWork.aspx#2>.
- Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*. Retrieved from <http://hbr.org/2003/12/the-one-number-you-need-to-grow/ar/1>.
- 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., Pearlman, K., & RothsteinHirsh, 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.

Schmit, M. J., & Allscheid, S. P. (1995). Employee attitudes and customer satisfaction: Making theoretical and empirical connections. *Personnel Psychology, 48*, 521-536.

Sechrest, L., & Yeaton, W. H. (1982). Magnitudes of experimental effects in social science research. *Evaluation Review, 6*(5), 579-600.

APPENDIX A. RELIABILITIES OF BUSINESS/WORK-UNIT OUTCOMES

Based on Schmidt & Hunter, 1996, Scenario 23, p. 219

Revenue/Sales		Profitability		Attrition	
Reliability	Frequency	Reliability	Frequency	Reliability	Frequency
1.00	4	1.00	3	1.00	1
0.99	2	0.99	2	0.63	1
0.92	2	0.94	1	0.62	1
0.90	1	0.93	1	0.60	1
0.62	1	0.91	1	0.39	1
0.57	1	0.90	1	0.27	1
0.34	1	0.89	2	0.25	1
		0.79	1	0.24	1
		0.57	1		
		0.56	1		

APPENDIX B. TEST-RETEST RELIABILITIES OF B2B CUSTOMER ENGAGEMENT

Based on Schmidt & Hunter, 1996, Scenario 23, p. 219

Customer Engagement	
Reliability	Frequency
0.30	1
0.78	1
0.44	1
0.11	1

APPENDIX C. BETWEEN-COMPANY ARTIFACT DISTRIBUTION OF RANGE-RESTRICTION/VARIATION ESTIMATES

Reliability					
s/S	Frequency	s/S	Frequency	s/S	Frequency
1.551	1	0.681	1	0.472	1
1.527	1	0.671	1	0.463	1
1.254	1	0.664	1	0.455	1
1.149	1	0.661	1	0.441	1
1.085	1	0.658	1	0.420	1
1.013	1	0.656	1	0.410	1
1.002	1	0.652	1	0.396	1
1.000	1	0.650	1	0.366	1
0.985	1	0.644	1	0.364	1
0.970	2	0.642	1	0.358	1
0.952	1	0.623	1	0.330	1
0.933	1	0.622	1	0.329	1
0.894	1	0.616	2	0.314	1
0.892	1	0.611	1	0.298	1
0.857	1	0.605	1	0.294	1
0.848	1	0.596	1	0.287	1
0.822	1	0.585	1	0.275	1
0.803	1	0.575	1	0.259	1
0.793	1	0.572	1	0.191	1
0.773	1	0.564	1	0.144	1
0.761	1	0.556	1		
0.755	1	0.537	1		
0.752	1	0.531	1		
0.750	1	0.517	1		
0.743	1	0.515	1		
0.740	1	0.513	1		
0.733	1	0.512	1		
0.717	1	0.508	1		
0.700	1	0.483	1		
0.696	1	0.475	1		

Values less than 1.000 indicate range restriction; values greater than 1.000 indicate range enhancement and produce a downward correction in the observed correlation. s=within-company standard deviation; S=standard deviation in the data pooled across companies; s/S=range variation ratio.

APPENDIX D. WITHIN-COMPANY ARTIFACT DISTRIBUTION OF RANGE-RESTRICTION/VARIATION ESTIMATES

Reliability			
s/\bar{S}	Frequency	s/\bar{S}	Frequency
2.47	1	1.11	1
2.43	1	1.09	1
2.00	1	1.07	1
1.73	1	1.06	1
1.62	1	1.05	3
1.60	1	1.04	2
1.59	1	1.03	1
1.57	1	1.02	1
1.55	2	0.99	2
1.52	1	0.98	2
1.49	1	0.97	1
1.43	1	0.96	1
1.42	1	0.95	1
1.37	1	0.93	1
1.35	1	0.92	1
1.31	1	0.91	1
1.28	1	0.90	2
1.26	1	0.86	1
1.23	1	0.85	1
1.21	1	0.82	4
1.20	3	0.81	1
1.18	2	0.77	1
1.17	1	0.76	1
1.14	1	0.75	1
1.12	1	0.74	1
0.73	1	0.47	2
0.70	1	0.46	1

Reliability			
s/\bar{S}	Frequency	s/\bar{S}	Frequency
0.67	1	0.44	1
0.65	1	0.41	1
0.63	1	0.30	1
0.58	2	0.23	1
0.57	1	0.18	1
0.53	1	0.16	1
0.52	1	0.14	1
0.50	1	0.08	1

Values less than 1.00 indicate range restriction; values greater than 1.00 indicate range enhancement and produce a downward correction in the observed correlation. s =within-company standard deviation; s/\bar{S} =average standard deviation across companies; s/\bar{r} =range variation ratio.

GALLUP®

World Headquarters

The Gallup Building
901 F Street, NW
Washington, D.C. 20004

t +1.877.242.5587

f +1.202.715.3045

www.gallup.com