Pierre Chandon, Vicki G. Morwitz, & Werner J. Reinartz

Do Intentions Really Predict Behavior? Self-Generated Validity Effects in Survey Research

Studies of the relationship between purchase intentions and purchase behavior have ignored the possibility that the very act of measurement may inflate the association between intentions and behavior, a phenomenon called "self-generated validity." In this research, the authors develop a latent model of the reactive effects of measurement that is applicable to intentions, attitude, or satisfaction data, and they show that this model can be estimated with a two-stage procedure. In the first stage, the authors use data from surveyed consumers to predict the presurvey latent purchase intentions of both surveyed and nonsurveyed consumers. In the second stage, they compare the strength of the association between the presurvey latent intentions and the postsurvey behavior across both groups. The authors find large and reliable self-generated validity effects across three diverse large-scale field studies. On average, the correlation between latent intentions and purchase behavior is 58% greater among surveyed consumers than it is among similar nonsurveyed consumers. One study also shows that the reactive effect of the measurement of purchase intentions is entirely mediated by self-generated validity and not by social norms, intention modification, or other measurement effects that are independent of presurvey latent intentions.

Consumers' self-reported intentions have been used widely in academic and commercial research because they represent easy-to-collect proxies of behavior. For example, most academic studies of satisfaction use consumers' intentions to repurchase as the criterion variable (for an exception, see Bolton 1998), and most companies rely on consumers' purchase intentions to forecast their adoption of new products or the repeat purchase of existing ones (Jamieson and Bass 1989). However, it is well known that consumers' self-reported purchase intentions do not perfectly predict their future purchase behavior, nor do these differences cancel each other out when intentions and behavior are aggregated across consumers. In a meta-analysis of 87 behaviors, Sheppard, Hartwick, and Warshaw (1988) find a frequency-weighted average correlation between intentions and behavior of .53, with wide variations across measures of intentions and types of behavior (for a review, see Morwitz 2001).

Pierre Chandon is Assistant Professor of Marketing, INSEAD, and currently he is Visiting Assistant Professor of Marketing, Kellogg School of Management, Northwestern University (e-mail: pierre.chandon@insead.edu). Vicki G. Morwitz is Associate Professor of Marketing and Robert Stansky Faculty Research Fellow, Stern School of Business, New York University (e-mail: vmorwitz@stern.nyu.edu). Werner J. Reinartz is Associate Professor of Marketing, INSEAD (e-mail: werner.reinartz@insead.edu). The authors acknowledge the helpful input of the anonymous *JM* reviewers, John Lynch, Gilles Laurent, and Albert Bemmaor, as well as those who participated when the authors presented this research at INSEAD; at the Association for Consumer Research Conference, Portland; at the Marketing Science Conference, College Park, Md.; and at the ESSEC-HEC-INSEAD conference. In addition, the authors thank the French online grocer Prodigy Services Company and Allison Fisher for providing data and INSEAD for its financial assistance.

To improve the ability to forecast behavior from intentions, researchers have tested alternative scales (Reichheld 2003; Wansink and Ray 2000) and have developed models that account for biases in the measurement and reporting of intentions, the heterogeneity across customers, changes in true intentions between the time of the survey and the time of the behavior, and the stochastic and nonlinear nature of the relationship between intentions and behavior (Bemmaor 1995; Hsiao, Sun, and Morwitz 2002; Juster 1966; Kalwani and Silk 1982; Manski 1990; Mittal and Kamakura 2001; Morrison 1979). In practice, the studies adjust the intention scores by analyzing the actual purchase behavior of consumers whose purchase intentions have been measured previously. For example, the popular ACNielsen BASES model forecasts aggregate purchase rates by applying conversion rates to measured purchase intentions (e.g., it assumes that 75% of consumers who checked the top purchase-intentions box will actually purchase the product). To obtain these conversion rates, BASES uses previous studies that measured the purchase intentions of consumers and then tracked their actual purchases.

However, a limitation of these studies is that they focus on the internal rather than the external accuracy of purchase-intention measures. That is, the studies measure the improvement in the ability to forecast the behavior of consumers whose intentions they previously measured, not the behavior of consumers whose intentions they did not measure. Therefore, the studies assume that they can extrapolate the intention–behavior relationship of nonsurveyed consumers on the basis of the relationship that surveyed consumers exhibit. In doing so, the studies ignore the potentially important problem that the measurement of intentions itself might self-generate some of the association between the intentions and the behavior of a particular consumer (Feldman and Lynch 1988).

Finding that part of the predictive power of purchase intentions is an artifact of the measurement would have serious implications for researchers and managers. It would suggest that studies that measure the strength of the association between intentions and behavior on the same sample of consumers overstate the external predictive accuracy of purchase intentions. This would explain why so many new products fail even after they perform well in purchase-intention tests. In general, researchers who are interested in measuring the true association between two constructs (in this case, for consumers whose behavior was not influenced by the measurement) would need a method that detects and corrects for the effects of measurement.

In this research, we develop a comprehensive latent framework to conceptualize the reactive effects of the measurement of purchase intentions. This framework distinguishes between two sources of measurement reactivity. The first is self-generated validity effects, which we define as a strengthened relationship between latent intentions and behavior due to the measurement of intentions. The second source includes all measurement effects that are independent of latent intentions, such as those that social norms or postsurvey intention modifications create.

We also describe a two-stage procedure to detect whether the act of measurement alters the strength of the relationship between a latent construct that is measured through surveys, experiments, or observations and its consequence (e.g., intentions–behavior, attitudes–intentions, attitudes–behavior, satisfaction–behavior) and to determine the true magnitude of the relationship in the absence of measurement. We demonstrate three empirical applications of this method using large-scale data sets that contain purchase or profitability data from both consumers whose purchase intentions were measured and similar consumers whose purchase intentions were not measured. In the three applications (groceries, automobiles, and personal computers [PCs]), we show that the strength of the relationship between latent intentions and behavior is stronger for surveyed consumers than for similar nonsurveyed consumers. In the final section, we discuss the managerial and research implications of our results.

Self-Generated Validity and Other Sources of Measurement Reactivity

Reactive Effects of Measurement

Ample evidence indicates that measurement can influence both the intensity of a measured construct and its association with other constructs. In intentions research, the reactive effects of measurement have been called the "mere measurement effect," "the self-erasing error of prediction," and "self-prophecy." We refer to the behavioral differences between surveyed and nonsurveyed consumers as the "reactive effects of measurement" or simply as "measurement reactivity."

In competitive markets in which most existing customers have positive attitudes toward a product category, the measurement of purchase intentions increases purchasing in the category of accessible and preferred brands. Research has shown these effects for both hypothetical and real brands, for financially important and relatively inconsequential behaviors, and for short (a few minutes) and long (six months) delays between the measurement and the behavior (Chandon, Morwitz, and Reinartz 2004; Dholakia and Morwitz 2002b; Fitzsimons and Morwitz 1996; Morwitz and Fitzsimons 2004).

In a related stream of research, studies show that asking consumers to predict their future behavior influences the likelihood that they will engage in that behavior (Sherman 1980; Spangenberg 1997; Spangenberg and Greenwald 1999; Sprott et al. 2003). Focusing on socially normative behavior, these studies demonstrate that if respondents are asked to predict the likelihood that they will perform a behavior in the future, they are more likely to engage in socially desirable behaviors, such as voting or recycling, and less likely to engage in socially undesirable behaviors, such as singing "The Star-Spangled Banner" over the telephone.

Self-Generated Validity Theory

The self-generated validity theory (Feldman and Lynch 1988), the most popular explanation of the reactive effects of measurement, uses two lines of argument. First, preexisting intentions may become more accessible in memory when the researcher asks the question. (It is also possible that consumers have no preexisting intentions and form them only in response to the researcher's question.) The measurement process thereby leads survey respondents to form judgments that they otherwise would not access in their memory or that they otherwise would not form. Second, higher relative accessibility and diagnosticity of intentions, compared with other inputs for purchase decisions (e.g., tastes, mood, competitive environment), may make subsequent purchase behavior more consistent with prior intentions.

Several studies provide indirect evidence in support of the self-generated validity theory for public opinion (Simmons, Bickart, and Lynch 1993) and marketing research (Fitzsimons and Morwitz 1996; Morwitz and Fitzsimons 2004; Morwitz, Johnson, and Schmittlein 1993). However, none has examined the core prediction of self-generated validity theory directly, namely, that the association between prior intentions and behavior is stronger among surveyed consumers than among similar nonsurveyed consumers. The studies have been unable to test this prediction because they have not estimated the purchase intentions of consumers who were not surveyed.

Consistent with Feldman and Lynch's (1988) predictions, Fitzsimons and Morwitz (1996) find that the measurement of general intentions to purchase automobiles increases the likelihood that buyers will repurchase the automobile brand that they previously owned and that first-time buyers will purchase brands with large market shares. Under the assumption that prior purchase rates or market shares are proxies for latent, brand-specific purchase intentions, Fitzsimons and Morwitz's results suggest that the measurement of general intentions increases the association between latent, brand-specific intent and brand choice.

Similarly, in a series of laboratory studies, Morwitz and Fitzsimons (2004) find that the measurement of general purchase intentions for candy bars makes consumers more likely to choose brands that they like and less likely to choose those that they dislike. If we consider prior brand preference a proxy for intention to purchase the brand, this study suggests that the measurement of consumers' intentions to buy from the category increases the association between their latent intentions to buy the brand and the likelihood of subsequently choosing this brand. Finally, Morwitz, Johnson, and Schmittlein (1993) examine the effects of repeated measurements of intentions (and behavior) on people with high and low initial measured purchase intentions. They find that the repeated measurement of intentions and behavior increases the association between behavior and the initial measure of intentions. However, their analysis is restricted to consumers whose purchase intentions have been measured at least once.

Other Sources of Measurement Reactivity

Ignoring obvious alternative explanations, such as selection biases, that violate our definitional assumption that surveyed and nonsurveyed consumers are identical, we consider at least two other explanations for the reactive effects of measurement in purchase-intentions surveys: social norms and intention modification. Both differ from self-generated validity in that they operate independently of consumers' intentions at the time of the survey.

In the context of socially normative behaviors, Sherman (1980) shows that asking people to predict their future behavior biases their reported intentions toward a social norm (e.g., donating to charities, not singing over the telephone). Consumers then act according to their newly reported intentions, not according to their prior unreported intentions, to reduce the cognitive dissonance between their reported intentions and their behavior (Spangenberg and Greenwald 1999; Sprott et al. 2003).

With regard to intention modification, consumers tend to evaluate market research surveys positively because they either find the survey informative or enjoy being asked their opinion (Dholakia and Morwitz 2002a; Sudman and Wansink 2002). In a subsequent stage, this positive evaluation of the survey carries over to the evaluation of the company and its products. Consumers also regard the survey as a signal of the firm's customer orientation, which directly improves their evaluation of the company and its products. In both cases, the positive attitude triggered by the survey leads to greater purchasing by surveyed consumers.

Both explanations share the view that the measurement of purchase intentions modifies consumers' purchase intentions rather than makes prior intentions more accessible in memory or more diagnostic of future purchase decisions. In the context of purchase-intention surveys for common products and services, the measurement effects make consumers more likely to report positive purchase intentions and then actually purchase the product, regardless of their purchase intentions at the time of the survey.

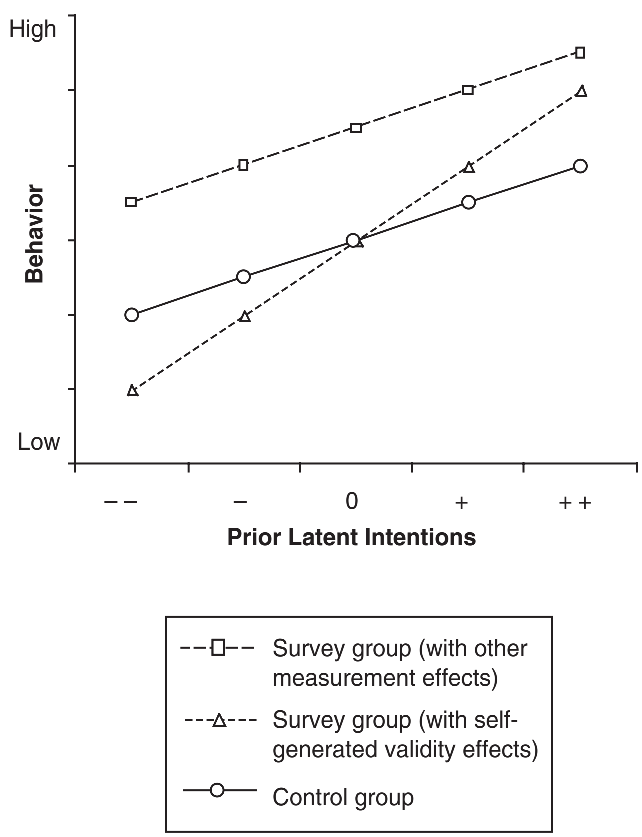
Summary

To better understand the differences between the possible sources of measurement reactivity, in Figure 1 we plot hypothetical purchase behavior (e.g., purchase quantity) as a function of presurvey, latent (i.e., unmeasured) purchase intentions for both consumers whose intentions were not measured (control group) and those whose intentions were measured.

In Figure 1, we show that the different sources of measurement reactivity have markedly different effects on purchase behavior and on the link between intentions and behavior. Intention modification leads to a consistent upward shift in purchase behavior but leaves the slope of the relationship between presurvey intentions and behavior unchanged. In contrast, self-generated validity effects do not lead to a general increase in purchase behavior but strengthen the association between intentions and behavior. If measurement reactivity is due to self-generated validity, intention measurement makes consumers with positive purchase intentions more likely to purchase but also makes consumers with negative purchase intentions less likely to purchase, which increases the steepness of the slope between intentions and behavior.

In Figure 1, we also show that in contrast to intention modification, self-generated validity effects do not necessarily lead to measurement reactivity. For example, the

FIGURE 1
Self-Generated Validity and Other Sources of Measurement Reactivity



measurement of intentions does not change the purchase behavior of consumers who have neutral purchase intentions, that is, those who are undecided about purchasing and not purchasing. Similarly, self-generated validity effects cancel out if there are as many positively inclined consumers as there are negatively inclined ones (i.e., if the distribution of purchase intentions is symmetric around the neutral point). In this case, the average purchase behavior of surveyed consumers may be the same as the average purchase behavior of similar nonsurveyed consumers, though the purchase behavior of each consumer is more extreme. However, self-generated validity effects are a sufficient condition for measurement reactivity when the majority of consumers have positive purchase intentions—the most common case in field studies of actual products in competitive markets—because the measurement of purchase intentions makes these consumers more likely to follow their intentions (i.e., more likely to purchase).

Conceptualizing and Estimating the Reactive Effects of Measurement

A Latent Model of the Effects of the Measurement of Purchase Intentions

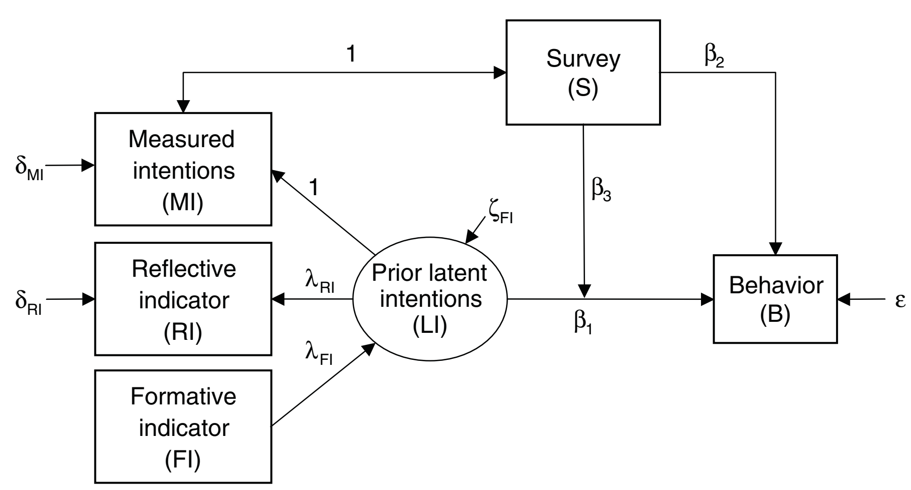
The framework we present in Figure 2 relates purchase behavior (B) to measured (self-reported) purchase intentions (MI), prior latent (unmeasured) purchase intentions (LI), and the survey that measures purchase intentions (S). In line with conventional representations of structural equation models, we use rectangles to represent observed variables, ovals for latent variables, arrows between constructs for causal relations, an arrow pointing to another arrow for an interaction effect, and a double arrow for a correlation.

We consider LI an unobserved hypothetical construct that captures, without error, consumers' determination to purchase just before the time of the survey. Thus, B is a function of LI (with regression coefficient β_1) and random error (ε). In the model, we assume that all consumers, both surveyed and nonsurveyed, have some latent purchase intentions at the time of the survey. However, this assumption does not imply that consumers have decided whether to buy before the survey, because prior latent intentions can be neutral; rather, it implies that consumers do not form intentions only when they are surveyed (we explore the implications of this assumption in the "General Discussion" section). By definition, these prior latent intentions are independent of whether the consumers' intentions are surveyed or not. If S is randomly administered, LI are identical for surveyed and nonsurveyed consumers, as we show in Figure 2 by excluding a link between S and LI.

We present the observed measures of LI on the left-hand side of Figure 2. Purchase intentions measured by the survey constitute one such measure, but this is not the only one. We can also measure latent intentions by other reflective indicators (denoted RI_1, RI_2, …, RI_n), including indirect measures, such as physiological measures or implicit tests, and behavioral measures, such as information search or the purchase of complementary products. Both LI and the measurement error (δ_{RI}) influence these reflective indicators. Other indicators of LI may be formative (e.g., prior purchase behavior, demographics), in which case LI is a function of the m formative indicators (denoted FI_1, FI_2, …, FI_m) and a random disturbance term (ζ_{FI}). We assume that these other indicators are independent of intention measurement (no correlation with S), whereas MI exist only for surveyed consumers (the correlation between MI and S is one). To identify the latent model, we must scale it by choosing one indicator for which the factor loading is set to one and the intercept is zero. Choosing MI as the scaling indicator enables us to scale the LI to the familiar units of MI. In doing so, we assume that there are no systematic reporting biases and that surveyed consumers retrieve their prior LI from memory. (We subsequently report simulation studies in which we examine what happens if MI are systematically biased upward because of social norms or intention modification.)

With the latent model, we can define self-generated validity effects more broadly. Originally, Feldman and Lynch (1988) studied the effects of measurement on the

FIGURE 2
A Latent Model of the Reactive Effects of the Measurement of Purchase Intentions



observed correlations among constructs. For example, Simmons, Bickart, and Lynch (1993) asked specific questions about the strength of election candidates before or after they measured general voting intentions. They then measured the impact of question order on the observed correlation between answers to specific questions and general voting intentions. We argue that the measurement of intentions makes presurvey latent intentions relatively more accessible and diagnostic than it does other antecedents of behavior, which strengthens the relationship between presurvey latent intentions and postsurvey behavior. Therefore, we represent self-generated validity in Figure 2 by the β_3 parameter, or the effect that S has on the link between LI and B.

This broader definition enables us to test for self-generated validity effects among latent (nonmeasured) and observed (measured) constructs and not only between observed constructs, as in Simmons, Bickart, and Lynch's (1993) study. It also excludes social norms and intention-modification effects, both of which imply that the surveying of intentions increases purchase behavior independent of prior latent intentions and that the relationship between prior latent intentions and behavior remains the same. However, these other sources of measurement reactivity lead to an increase in purchase behavior, regardless of prior latent intentions. Therefore, in Figure 2, we represent their effects by the β_2 parameter, which captures the effect of S on B and is not mediated by the strengthening of the relationship between LI and B.

Model Estimation

The right-hand side of the latent model in Figure 2 can be expressed as the following latent equation:

$$(1)\qquad B = \alpha_1 + \beta_1(LI) + \beta_2(S) + \beta_3(LI)(S) + \varepsilon,$$

where B is the future purchase behavior of interest; LI is latent purchase intentions; S is a binary variable that indicates whether intentions are surveyed; ε is the error term that captures random disturbance; and α_1, β_1, β_2, and β_3 are parameters to be estimated.

Parameter interpretation. In Equation 1, the β_3 parameter of the interaction between S and LI on B captures self-generated validity effects. When S is coded as .5 for surveyed consumers and –.5 for nonsurveyed consumers, we expect β_3 to be positive, which indicates a higher association between LI and B for surveyed consumers than for similar nonsurveyed consumers.

As Irwin and McClelland (2001) explain, the β_1 and β_2 coefficients of the LI and S variables capture the simple effects of each variable when the other variable involved in the interaction is zero. Therefore, the β_1 parameter captures the mean effect of LI on B across both surveyed and nonsurveyed consumers. Because LI are scaled according to MI, the interpretation of the β_2 parameter depends on whether MI are measured on a bipolar or a unipolar scale.

When purchase intentions are mean-centered and measured on a unipolar scale (e.g., a timed intent scale ranging from "intend to buy immediately" to "will never buy"), the β_2 parameter captures the effects that measurement has for consumers with average LI. However, when purchase intentions are measured on a bipolar interval scale with a neutral point (e.g., 3 on a five-point scale, where 1 = "completely disagree" and 5 = "completely agree") and are centered on this neutral midpoint, β_2 captures the effect of measurement on the purchase behavior of consumers with neutral purchase intentions. In other words, β_2 measures the sources of measurement reactivity that are due not to self-generated validity but rather to social norms or intention modification. (This is because making a neutral purchase intention more accessible or more diagnostic should not influence behavior.) Any differences in purchase behavior between surveyed and nonsurveyed consumers with a neutral intent to buy cannot be explained by self-generated validity effects and therefore must be attributable to these other explanations.

This interpretation of β_2 requires a set of assumptions. First, the construct of interest is valenced (i.e., can be positive, negative, or neutral), which is not problematic if the construct of interest is attitude or satisfaction, both of which are valenced constructs. Many studies of purchase intentions also assume that intentions are valenced, at least implicitly (e.g., when the studies measure intentions on a bipolar Likert scale). This assumption is inconsistent with Fishbein and Ajzen's (1975) definition of behavioral intentions as a probability, or a unipolar concept. Second, all consumers view answering at the midpoint of a valenced scale (e.g., "neither agree nor disagree") as a neutral, nonvalenced intention (i.e., consumer heterogeneity with respect to this perception is evenly distributed). This assumption is problematic, for example, in cross-national research in which there should be strong differences in response styles across countries. However, note that none of these assumptions is required to interpret β_3, the main coefficient of interest, which captures self-generated validity effects regardless of whether the intentions are measured on a bipolar or a unipolar scale.

Two-stage estimation. The difficulty of estimating Equation 1 is that LI is an unobserved latent variable, but fortunately we can estimate such a model using a two-stage approach (Bollen 1996; Bollen and Paxton 1998). As we detail in the Appendix, we can substitute LI in Equation 1 with MI – δ_{MI} to obtain an equation with only observed variables. Because MI is correlated with the new composite disturbance term (μ), which now includes δ_{MI}, ordinary least squares (OLS) cannot estimate the modified Equation 1. In addition, MI is missing for the control group of consumers who did not answer the survey.

To overcome these obstacles, we regress MI on the other indicators of LI (RI_1, RI_2, ..., RI_n; FI_1, FI_2, ..., FI_m) using data from the survey group. These other indicators serve as instrument variables for MI because they are correlated with LI but not with δ_{MI} (and therefore not with the new composite disturbance term in Equation 1). We then use the fitted parameters of this regression to substitute MI into Equation 1 with its predicted value $\hat{MI}$ in both the survey and the control groups. Because $\hat{MI}$ is a linear combination of variables that are not correlated with μ, $\hat{MI}$ is not correlated with μ. Thus, we can use an OLS regression to estimate the modified equation, including $\hat{MI}$.

Simulation analyses. To estimate the model in Figure 2 and Equation 1, we must assume that (1) multiple indicators of LI are available, (2) MI are unbiased indicators of LI that are unaffected by social norms or intention modification, and (3) surveyed and nonsurveyed consumers are identical (i.e., there are no selection biases). We tested the importance of each assumption by conducting extensive simulations, which also enabled us to estimate the ability of the two-stage procedure to recover the true effects of LI on B for surveyed and nonsurveyed consumers when these assumptions were not satisfied. Specifically, we manipulated the quality of the other indicators of LI (factor loadings ranging from .3 to .9), the presence of positive reporting biases in MI (e.g., those caused by social norms or intentions modification), and the presence of selection biases (only positively inclined consumers agree to answer the survey). We find that the β_3 coefficients estimated with the two-stage procedure are stable even in the extreme scenario of reporting or selection biases combined with poor indicators of LI. In addition, these problems inflate the standard errors of the coefficients and work against our hypotheses. Overall, the simulation analyses show that the two-stage procedure can estimate self-generated validity effects reliably even in imperfect measurement and experimental conditions. (The complete results of the analyses are available from the authors on request.)

Summary

The latent model enables us to broaden our definition of self-generated validity effects to include the effects that measurement has on the relationship between latent and measured constructs. We show that with a two-stage procedure, we can estimate (1) the latent purchase intentions of nonsurveyed consumers, (2) the impact of the measurement of intentions on the relationship between latent intentions and behavior (self-generated validity effects), and (3) the impact of the measurement of intentions on the behavior of consumers with neutral intentions (social norms and intention-modification effects).

Quantifying Self-Generated Validity and Other Sources of Measurement Reactivity: Three Field Studies

In this section, we quantify self-generated validity and other sources of measurement reactivity through three large-scale studies of intended and actual purchases of groceries, automobiles, and PCs. The three field studies differ significantly in terms of the sampling frame, the type of purchase behavior studied, and the measurement of intentions and behavior, but they all contain information about purchasing from two groups of consumers: those whose purchase intentions were measured and a control group of similar consumers whose purchased intentions were not measured. Therefore, we describe the three studies and their results collectively.

Data

Grocery study. In this study, we measured consumers' intentions to repurchase from an online grocer. The data (for a detailed description, see Chandon, Morwitz, and Reinartz 2004) were gathered from a field study conducted in collaboration with a leading French Web-based grocer that offers online an assortment that is typical of a large supermarket (50,000 stockkeeping units of food and some durable products) and nationwide delivery. During the last week in May and the first week in June 2002, 251 customers were contacted by telephone and asked about their intent to repurchase from the online grocer in the future. The respondents were chosen at random from customers who had made their first purchase with the online grocer in October or November 2001. The data set contained demographic information about the age, number of children, and number of pets of each customer, as well as detailed transaction data for all their purchases between January 2001 and April 2002 (i.e., nine months before and nine months after the survey). Transaction data included the date of the order, the quantity and price of each ordered product, and the total profit that the online grocer made. The same data were available for a control group of 140 consumers who were randomly selected from the same cohort but whose purchase intentions were not measured.

To obtain reliable indicators of purchase intentions, we measured consumers' agreement with two statements (translated from French): "I am thinking of using [name of grocery company] for my next online purchases," and "I am thinking of remaining a customer of [name of grocery company] for a long time." We measured the statements on a five-point Likert scale, where 1 = "completely disagree" and 5 = "completely agree." We averaged the answers to produce a reliable intention scale (Cronbach's α = .86). Because we surveyed existing buyers, in general the intention scores were above the midpoint (mean = 3.86, standard error [S.E.] = 1.19, $p < .01$) and somewhat negatively skewed (skewness = –1.13). However, there was enough variance in the measures to test for self-generated validity effects (26% of consumers had negative to neutral [i.e., below 3] intentions to purchase).

For the grocery study, we examined two dependent variables: purchase incidence and customer profitability. We chose purchase incidence to facilitate comparisons with our other two studies and with previous measurement-reactivity research and to examine the effects of the measurement of intentions on the consumers' first repurchase after the survey. Using a binary variable, we determined whether the consumer placed at least one order halfway through the postsurvey period (just more than four months after the survey). We chose this time horizon because self-generated validity theory predicts, and prior research demonstrates, that the effects of the measurement of purchase intentions decay over time (Chandon, Morwitz, and Reinartz 2004).

We studied cumulative customer profitability because it includes information about the first and subsequent repurchases that the customer made. In addition, it is the most important measure for managers, and prior research has shown that one-time transactional gains do not necessarily lead to improved customer profitability for the company, especially for commonly purchased consumer goods (Reinartz and Kumar 2000). We measured total customer profitability as the cumulative net profit attributable to the customer (i.e., the sum of the contribution of all orders placed in the nine-month postsurvey period less coupons and delivery costs). Because the company routinely surveys its customers and would continue to do so even in the absence of measurement-induced purchases, we treated the cost of administering the survey as a fixed cost and did not subtract it from the cumulative contribution of surveyed customers. We used the full postsurvey period because we measured cumulative profits, which include the first and subsequent purchases. Although the effects of measurement decay over time, the positive impact of the first purchase carries through to subsequent purchases; therefore, measurement-reactivity effects persist over time on cumulative customer profitability (Chandon, Morwitz, and Reinartz 2004).

Automobile and PC studies. The other two data sets refer to the automobile and PC data that Morwitz, Johnson, and Schmittlein (1993) use and describe. Intentions to buy and ownership of home PCs and automobiles were measured using two different but similar U.S. consumer mail panels. Both panels were designed to be representative of U.S. households, according to census data, and each panel comprised approximately 100,000 households. Intentions and behavior were measured during seven survey waves, each approximately six months apart. The surveys requested that the person in the household who was most involved in the purchase decision complete the survey. In each survey wave, panel households were asked to provide their timed intentions for buying an automobile or PC in the future. Extensive demographic information about the panel households also was collected, including the size of the household, annual household income, age of the head of the household, marital status, home ownership, household stage of life, occupation, education of the head of the household, race, number of cars owned, and regional dummy variables.

In contrast to the grocery data, for which consumers were randomly selected to be surveyed, Morwitz, Johnson, and Schmittlein's (1993) data for both products reflect the results of naturally occurring or quasi experiments. Because of panel dynamics (members entering and exiting a panel over time), panel members varied in whether and how often their intentions were measured. For both products, we compared the behavior of panel members who entered the panel only in time to receive the intentions question in the sixth survey wave with the behavior of those who joined the panel after the sixth but before the seventh wave and thus whose intentions were not measured in the sixth wave. As in Morwitz, Johnson, and Schmittlein's research, to control for any differences in the experimental and control groups due to factors other than the experiment, we weighted the data by two different criteria: stage in the life cycle and age of the head of the household. Because the results for both weighting schemes are similar, we report only the results for weighting by life cycle.

For the automobile data, the intention question asked, "When will the next new (not used) car (not truck or van) be purchased by someone in your household?" The following response alternatives were provided: 1 = "6 months or less," 2 = "7–12 months," 3 = "13–24 months," 4 = "25–36 months," 5 = "over 36 months," and 6 = "never." We reverse coded the responses so that higher numbers represent a higher intention to repurchase, which is consistent with the grocery data (mean = 2.36, S.E. = .027, $p < .01$, skewness = –.74). During each survey wave, respondents also were asked whether they had purchased a new automobile during the previous period. For the automobile study, we analyzed data from 8347 households, 3571 whose intentions were measured and 4776 whose intentions were not measured.

The PC data are similar in format to the automobile data. The intention question asked, "Do you or does anyone in your household plan to acquire a (another) personal computer in the future for use at home?" The following response alternatives were provided: 1 = "yes, in the next 6 months"; 2 = "yes, in 7 to 12 months"; 3 = "yes, in 13 to 24 months"; 4 = "yes, sometime, but not within 24 months"; 5 = "no, but have considered acquiring one"; and 6 = "no, will not acquire one." We also reverse coded these responses (mean = 2.03, S.E. = .028, $p < .01$, skewness = –1.47). In each wave, respondents indicated whether they had purchased a computer in a given time period. As do Morwitz, Johnson, and Schmittlein (1993), we restricted our analysis to households that initially did not own a PC, and we assumed that a household bought a PC if it switched from being a nonowner to being an owner from one wave to the next. There were 7772 households in the data, 2138 whose intentions were measured and 5634 whose intentions were not measured.

For both the PC and the automobile data, we examined only purchase incidence (i.e., whether a purchase occurred in the six-month period following the intent measurement). Because intentions were measured during every survey wave, a longer-term analysis would confound duration with the number of times intentions were measured.

First-Stage Regressions

Predicting purchase intentions in the control group. In all three studies, we used demographic and behavioral indicators of LI as the instrument variables to predict MI in both the survey and the control groups. To select the instrument variables, we measured their predictive power by splitting the survey group into two random samples, regressing MI on the instrument variables, and then using the regression to predict intent in the second sample. As Armstrong and Collopy (1992) recommend, for both random samples, we selected the combination of variables with the best predictive accuracy and measured it with the median average percentage error (MdAPE) and the median relative absolute error (MdRAE). We obtained the MdRAE by dividing the median of the absolute forecast error by the corresponding error for the naive model, so we assigned the average MI of the survey group to all consumers in the control group. We then reestimated the best model of the MI for the full sample of consumers in the survey group and used the parameters from the regression to predict purchase intentions for both the survey and the control samples. To check the robustness of the final results for the choice and quality of the instrument variables, we tried several different predictions of intent that provided similar to significantly worse predictive power. The results were virtually unchanged.[1]

---

[1]We computed predicted purchase intentions for the grocery study as follows: $\hat{MI}$ = 5.467 – .00321(REC) – .03067(AGE) +

Across the three studies, we can predict purchase intentions moderately well. For the grocery data, MdAPE = .16 and MdRAE = .95; for the PC data, MdAPE = .087 and MdRAE = .78; and for the automobile data, MdAPE = .17 and MdRAE = .54. For both the PC and the automobile data, the error rates were similar when we used weighting methods to ensure equivalence between the survey and the control groups.

Method checks and descriptive results. As we expected, predicted purchase intentions were similar for surveyed and nonsurveyed consumers across all three studies (see Table 1). The difference between the groups was not statistically significant for the grocery study but was statistically significant for the automobile and PC studies; this is probably due to the larger number of observations in the latter studies (n = 8306 for the automobile study, n = 7772 for the PC study). The finding that predicted that purchase intentions are lower in the survey group than in the control group (as in two of three cases) helps rule out selection biases, which would cause consumers with higher LI to be more likely to appear in the survey group. Overall, the results indicate that surveyed and nonsurveyed consumers are similar and that the measurement of their purchase intentions causes the differences between their purchase behavior.

As Table 1 and previous studies with these data (Chandon, Morwitz, and Reinartz 2004; Morwitz, Johnson, and Schmittlein 1993) show, in general the measurement of purchase intentions increases the purchase incidence for grocery, automobile, and PC products. (Note that the PC data are marginally significant on the basis of a one-tailed test, as Morwitz, Johnson, and Schmittlein [1993] report.) In Table 1, we show that the variance in future behavior is lower in the control groups than in the survey groups for all studies and dependent variables (though the difference is not statistically significant in the PC study). This new result, which previous measurement-reactivity studies have not explored, is consistent with the self-generated validity theory, which argues that the measurement of a person's intention to perform a behavior, not other factors (e.g., mood, price promotions), increases the likelihood that he or she will act on this behavior. If we assume that the other factors cancel out, self-generated validity theory predicts that the purchase behavior of people in the control group will regress to the mean, which explains the lower variance in the control group. Alternatively, purchase intentions may cause more variance in behavior when they are made salient by measurement because they magnify the differences between people with high and low intentions. As we report subsequently, the variance differences have implications for the selection of the correct method to test for self-generated validity effects.

Second-Stage Regressions

Variable coding. We used the predicted purchase intentions that we estimated in the first stage to estimate the regression represented in Equation 1 in accordance with the procedure we detail in the Appendix. For the automobile and PC studies, whose timed measures of purchase intentions have no neutral point, we mean-centered the predicted purchase intentions ($\hat{MI}$) to zero when they measure 2.36 (automobile) and 2.04 (PC). However, for the grocery study, we measured purchase intentions on a five-point Likert scale (1 = "completely disagree" and 5 = "completely agree"), and therefore the neutral midpoint is 3 ("neither agree nor disagree"). To estimate the effects of the measure-

.34831(PET) – .40852(BABY), where REC is purchase recency (number of days since the latest purchase), AGE is the customer's age, and BABY and PET are dummy variables that indicate that the household has at least one child or one pet. For the automobile study, $\hat{MI}$ = 5.56050 – 1.38433(EDUCATION1) – 1.70868(EDUCATION2) – 1.38598(EDUCATION3) – 1.37272(EDUCATION4) – 1.20131(EDUCATION5) – 1.22879(EDUCATION6) – .00828(INCOME) – .46495(MANAGER) – .28947(TECH) + .09822(SERVICE) – .06099(FARM) + .06992(CRAFT) + .08894(OPERATOR), where EDUCATION1–6 are dummy variables that represent different levels of the education of the head of the household; INCOME is total household income; and MANAGER, TECH, SERVICE, FARM, CRAFT, and OPERATOR are dummy variables that represent different occupations of the head of the household. For the personal computer study, $\hat{MI}$ = 4.42389 + .01317(HH_AGE) – .28590(NEWHH) – .12844 (NEWBB) – .17102(LOWMIDF) – .62929(UPSCALE) – .04337(MIDAGE) + .29444(LOWMIDE), where HH_AGE is the age of the head of the household, and NEWHH, NEWBB, LOWMIDF, UPSCALE, MIDAGE, and LOWMIDE represent mutually exclusive levels of the customer life-cycle variable that the survey company created.

TABLE 1
Method Checks and Descriptive Results (Mean and Standard Deviation)

Study	Variable	Control Group	Survey Group
Grocery	Number of observations	140	251
	Predicted purchase intentions (1–5)	3.91 (.43)	3.86 (.41)
	Repeat purchase incidence	.229 (.421)	.331* (.471)*
	Customer profitability (€)	19.53 (42.27)	27.43 (57.08)**
Automobile	Number of observations	4776	3530
	Predicted purchase intentions (1–6)	2.25 (.768)	2.52** (.798)**
	Purchase incidence	.024 (.153)	.033* (.178)**
PC	Number of observations	5634	2138
	Predicted purchase intentions (1–6)	2.05 (.456)	2.02* (.468)**
	Purchase incidence	.038 (.191)	.045 (.207)

*p < .05.
**p < .01 (all tests two-tailed).
Notes: Standard deviations (in parentheses) are compared according to Levene's F-test.

ment of purchase intentions for consumers with neutral purchase intentions, we centered $\hat{MI}$ on the midpoint so that it equals zero when predicted purchase intentions measure 3. Finally, we coded the binary variable S, which captures the effect of intentions measurement, as .5 for consumers in the survey group and –.5 for consumers in the control group. We report the results (parameter estimates and standard error) of the second-stage regression in Table 2.

Model results. As we show in Table 2, consumers with higher LI are more likely to purchase in all three studies and are more profitable for the firm in the grocery study (the β_1 coefficients are all positive and statistically significant). Therefore, we replicate prior studies' findings that purchase intentions are a strong but imperfect predictor of purchasing. In addition, we find the expected interaction between latent intentions and intention measurement for all three studies and all dependent variables (the β_3 coefficients are all positive and statistically significant). Thus, LI are stronger predictors of the behavior of surveyed consumers than of nonsurveyed consumers. In other words, the measurement of purchase intentions strengthens the associations between latent purchase intentions and purchase behavior or customer profitability; this is a self-generated validity effect.

Finally, the β_2 coefficients, which capture the simple effects of the purchase intentions survey, are positive and statistically significant for the automobile and PC studies; thus, the measurement of purchase intentions increases future purchasing by consumers with average latent purchase intentions. Our two-stage method replicates previous findings from the same data that were obtained using different methods. However, for the grocery study, the β_2 coefficients for repeat purchase and customer profitability are not statistically different from zero, which demonstrates that the measurement of purchase intentions does not increase the purchases or profitability of consumers who have neutral latent purchase intentions.[2] Therefore, measurement reactivity in the grocery study is entirely mediated by self-generated validity effects.

Separate analyses for surveyed and nonsurveyed consumers. We performed the following analyses to obtain a more intuitive grasp of the magnitude of self-generated validity effects. We computed the correlation between predicted intentions and behavior in each group, for each study, and for each dependent variable. As we report in Figure 3, the results show that self-generated validity effects are great. On average, the correlation between intentions and behavior is 58% greater in the surveyed groups than in the control groups. In addition, the magnitude of the self-generated validity effects is approximately constant across all studies and dependent variables, regardless of the intensity of the true association between intentions and behavior (which varies between .07 in the automobile study and .26 in the grocery study).

To further illustrate the magnitude of self-generated validity effects, we regressed purchase behavior on predicted purchase intentions separately in the survey and control groups. As we show in Table 3, unstandardized regression coefficients for predicted purchase intentions are 76% greater in the survey groups than in the control groups. For example, a one-point difference in predicted purchase intentions (measured on a five-point scale) in the grocery study leads to a €52.71 gain in customer profitability when intentions are measured but only €23.95 when intentions are not measured. Similarly, although predicted purchase intentions are reliable predictors of purchase behaviors in all studies, the t-values are, on average, 70% greater in the survey groups than in the control groups. Taken together, the results show that the external accuracy of purchase intentions is significantly weaker and less reliable than is their internal accuracy and that we cannot extrapolate one from the other.

In Figures 4 and 5, we provide another perspective on the results by reporting the purchase behavior of three equal groups of consumers with low (<32nd percentile), moderate (33–66th percentile), and high (>67th percentile) predicted purchase intentions. In the grocery study (Figure 4), we find that the measurement of purchase intentions increases the repeat purchase incidence and customer profitability for high and moderate (positive) intenders but decreases both behaviors for low intenders, who have mostly negative intentions. The purchase-incidence findings from the automobile and PC studies (see Figure 5) replicate the same pattern. Overall, the pattern of results in Figures 4 and 5 matches the hypothetical self-generated validity effects

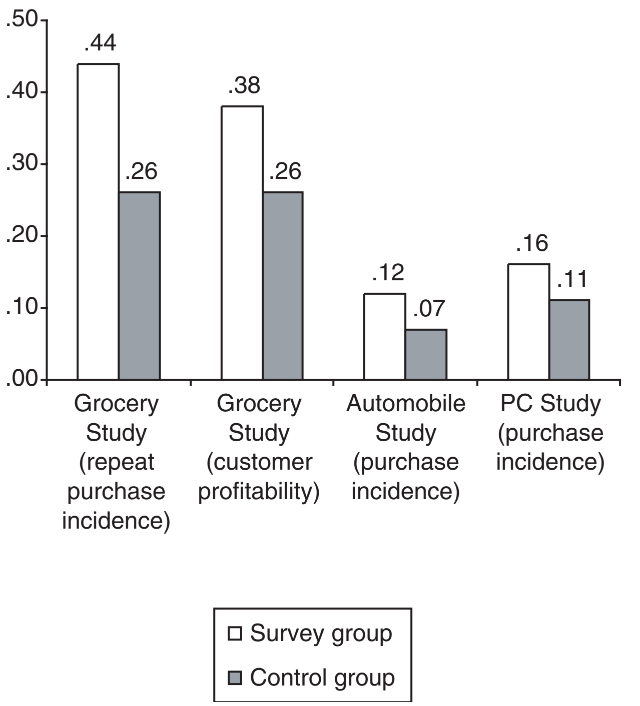
[2]If we mean-center predicted purchase intentions instead of centering them on the midpoint of the scale (3), the β_2 coefficients become positive and statistically significant for repeat purchase (β_2 = .12, S.E. = .04, $p < .01$) and, to a lesser extent, for customer profitability (β_2 = 9.73, S.E. = 5.16, $p < .10$). Using this coding, we replicate the descriptive results that we present in Table 1. The measurement of purchase intentions increases the purchasing and profitability of consumers who have average or, in this case, positive (mean = 3.86) purchase intentions, which is fully consistent with the predictions of self-generated validity theory.

TABLE 2
Output of Second-Stage Regression (Unstandardized Regression Coefficients and Standard Errors)

Study	Purchase Behavior	Prior Latent Intentions (β_1)	Survey (β_2)	Interaction (β_3)
Grocery	Repeat purchase incidence	.38** (.05)	–.11 (.10)	.26** (.11)
	Customer profitability	39.32** (6.20)	–13.70 (12.17)	26.78* (12.39)
Automobile	Purchase incidence	.02** (.00)	.06** (.02)	.01** (.00)
PC	Purchase incidence	.06** (.01)	.01* (.01)	.02* (.01)

*$p < .05$.
**$p < .01$ (all tests two-tailed).

FIGURE 3
Correlation Between Predicted Purchase Intentions and Behavior in Three Field Studies



shown in Figure 1 for all three studies and all dependent variables.

General Discussion

Because purchase intentions are widely used but are imperfect indicators of actual purchasing, a large body of research is devoted to improving their internal accuracy (the ability to predict the behavior of consumers from their previously measured intentions). However, we contribute to this literature by studying the external accuracy of measured intentions (i.e., their ability to predict the behavior of consumers whose intentions are not measured). We develop a comprehensive latent model of the reactive effects of the measurement of purchase intention in which we distinguish between two sources of measurement reactivity. The first is self-generated validity effects, which we define as a strengthened relationship between the measured latent construct and its behavioral consequences. Thus, self-generated validity effects increase the likelihood that consumers will follow their intentions. The theory behind these effects predicts that the measurement of intentions makes high intenders more likely to purchase and low intenders less likely to purchase but does not change the behavior of consumers with neutral intentions. The second source includes measurement effects that are independent of intentions, such as those created by social norms or intention modification. Unlike self-generated validity effects, the effect of social norms and intention modification influence the behavior of all consumers, regardless of their prior intentions.

We provide a two-stage procedure, which enables us to quantify the magnitude of the self-generated validity effects and other sources of measurement reactivity. In the first stage, we estimated the relationship between measured intentions and other indicators of latent intentions, using data from surveyed consumers. We then used the fitted parameters from our analysis to predict the latent purchase intentions of both surveyed and nonsurveyed consumers. In the second stage, we compared the strength of the association between our predicted intentions and actual behavior across both groups. Using data from three large-scale field studies with control groups, we find that the measurement of purchase intentions increases the association between latent intentions and purchase behavior. The effects are significant and robust across a variety of purchase behaviors, sampling frames, and ways to measure intentions and behavior. In addition, one study shows that the measurement of purchase intentions does not influence the purchases of consumers who have neutral purchase intentions, which suggests that self-generated validity effects cause all the reactive effects of measurement. The results have implications for both applied and academic research.

Managerial Implications

The obvious implication of our results is that commonplace procedures and models (e.g., ACNielsen's BASES model) that measure the intentions and behavior of the same sample of consumers overestimate the strength of their association. For most tested concepts, which elicit positive intentions in general, the models overstate aggregate purchase probabilities. Therefore, our results strongly call into question the common practice of extrapolating to the general population

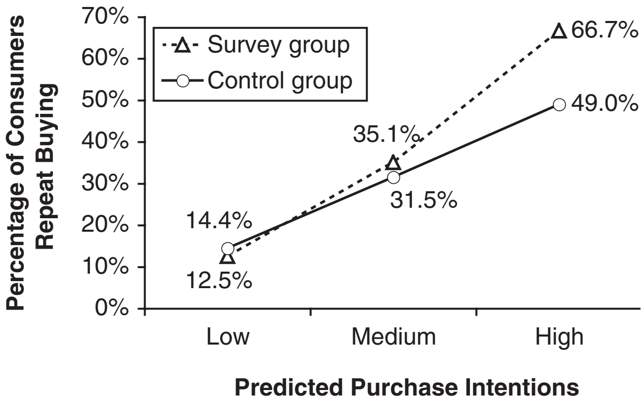
TABLE 3
Association Between Latent Purchase Intentions and Future Behavior in the Control and Survey Groups (Unstandardized Regression Coefficients and t-Values)

Study	Purchase Behavior	Control Group	Survey Group
Grocery	Repeat purchase incidence	.25 (3.11)	.51** (7.81)
	Customer profitability	25.93 (3.18)	52.71* (6.40)
Automobile	Purchase incidence	.015 (5.13)	.028** (7.39)
PC	Purchase incidence	.048 (8.71)	.069* (7.35)

*$p < .05$.
**$p < .01$ (all tests two-tailed).
Notes: Regression coefficients are compared with a Chow test (grocery: $F_{1,\ 389}$; automobile: $F_{1,\ 8302}$; PCs: $F_{1,\ 7768}$).

FIGURE 4
Surveyed and Nonsurveyed Consumers with Three Levels of Predicted Purchase Intentions (Grocery Study)

A: Repeat Purchase Probability



B: Customer Profitability

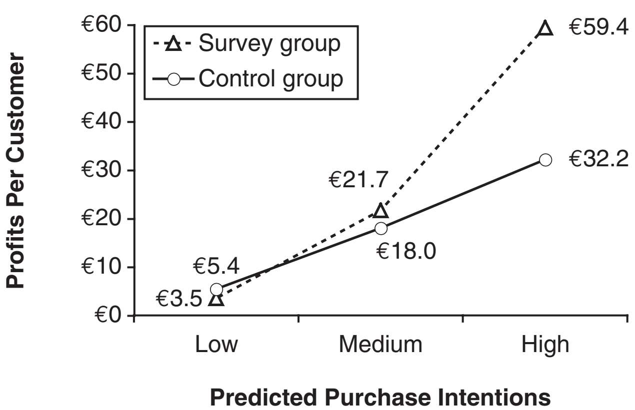
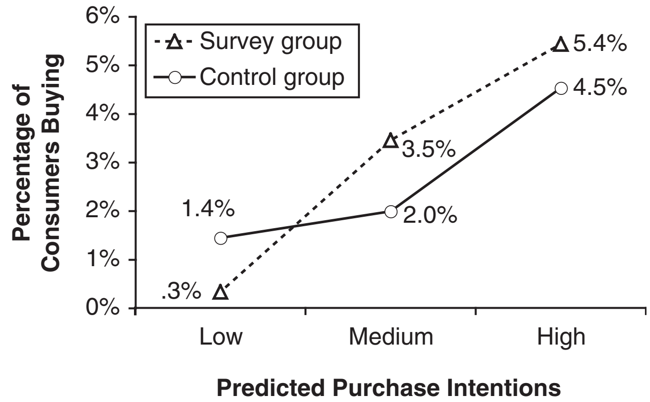
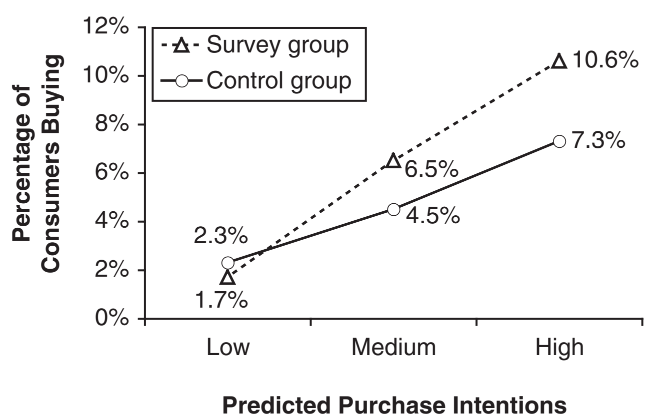


FIGURE 5
Purchase Probability of Surveyed and Nonsurveyed Consumers with Three Levels of Predicted Purchase Intentions

A: Automobile



B: PC



the conclusions of studies that measure the intentions and behaviors of the same consumers. When choosing the best measure of purchase intentions or the best method to predict behavior from intentions, marketers should focus on the external, not internal, validity of the measure and the method.

For example, there has been a recent debate in the *Harvard Business Review* about the merits of different measures of customer feedback. Reichheld (2003) argues in favor of measuring consumers' willingness to recommend the product, because he claims that it is the best predictor of future purchasing. However, this conclusion is based on a comparison of the predictive accuracy of different measures of customer feedback that are tested on surveyed consumers only. Our results suggest that marketers who are interested in selecting the measure that best predicts future purchasing should use our method to determine whether it predicts the behavior of nonsurveyed consumers.

Our results also emphasize the importance of investigating the sources of measurement-reactivity effects. Knowing whether self-generated validity or other measurement effects drive the behavioral differences between surveyed and nonsurveyed consumers has implications for the improvement of forecasting and targeting. For example, Jamieson and Bass (1989) describe multiple conversion schemes that marketers use to forecast purchase behavior from intentions. These conversion schemes are obtained by analyzing the behavior of consumers whose intentions have been measured. A scheme that Jamieson and Bass describe is 75%–25%–10%–5%–2% for each purchase-intention box (e.g., 75% of consumers who state that they would "definitely buy" actually do so, 25% of consumers who state that they would "probably buy" actually do so). If social norms or intention modification causes the reactive effects of measurement, these weights are inflated and should be reduced by a constant (e.g., the correct weighting scheme might be

60%–10%–0%–0%–0%). In this case, marketers should consider narrowing their target to focus only on consumers who have strong positive purchase intentions, because they are the only ones likely to purchase. However, if self-generated validity causes measurement reactivity, conversion rates should be regressed toward their means (e.g., the correct weighting scheme might be 60%–20%–15%–10%–8%). This flatter purchasing profile implies that marketers should broaden their target to encompass consumers who have negative purchase intentions because they are more likely to purchase than the conversion rates, which are determined by the measurement of surveyed consumers, suggest.

Research Implications

An area for further research is to relax the assumption that all consumers have some form of prior latent intention before the survey. For our research, it is reasonable to assume that an existing customer of a Web grocer has formed a repurchase intention or that a U.S. consumer would have an intention to buy an automobile or PC. However, as Feldman and Lynch (1988) argue, a segment of the population may form an intention only when asked about it. Although it would violate a main assumption of the model, we expect that our procedure would still be able to detect self-generated validity effects even if a sizable segment of consumers did not have a latent intention. Suppose the sample consists of a probability mixture of two groups, one lacking prior latent intention and one with latent intention. Then suppose that intention measurement causes a latent intention among respondents who did not previously have one and makes them more likely to follow this new intention. Finally, suppose that there is no change in the strength of the relationship between latent intentions and behavior for people in the group with preexisting latent intentions. The first stage of our two-stage procedure would incorrectly assign a purchase intention to the segment of consumers in the control group who have none. However, because the purchase behavior of these consumers remains independent of this predicted intention, in the second stage of the procedure, we would find that the association between the control group's latent intentions and behavior is small and different from the association between the intentions and behavior in the survey group. Thus, we believe that our estimation procedure is capable of detecting self-generated validity effects even in cases in which latent intentions do not exist before measurement.[3]

Our study relates to previous studies that have demonstrated that measurement-related biases can lead to incorrect inferences about the strength of the relationship between two measured marketing constructs and therefore have developed corrective techniques (Baumgartner and Steenkamp 1992; Rossi, Gilula, and Allenby 2001). For example, Greenleaf (1992) investigates how different response biases affect the relationship between self-reported attitudes and behavioral frequencies. For a large battery of behaviors, he develops a method to detect whether response styles reflect true attitude differences, in which case researchers should not adjust for them, or are biased, in which case researchers should adjust for them. A fruitful area for additional research would be to integrate the methods and findings from that stream of research with our method for the estimation of latent constructs among respondents for whom the constructs were not measured.[4]

The method we offer helps measure and correct for self-generated validity effects in many research contexts, including laboratory experiments and field observations, and for many constructs, including beliefs, attitudes, and satisfaction. When self-generated validity effects are possible, researchers should collect data about the criterion (e.g., behavior) of a control sample of consumers who did not answer the survey as well as multiple indirect measures of the target explanatory constructs (e.g., intentions, attitudes) that the survey does not influence (e.g., behavioral or demographic data that is measured with a different method than that used to measure the explanatory and criterion variables). With this information, researchers should be able to predict the level of the explanatory construct for a control group of consumers who did not answer the survey and to measure the link between the predicted and the criterion variables.

For example, this method could clarify inconsistencies between survey results and the behavior of the general population, such as in contingent valuation surveys for environmental policies or products (Irwin 1999). It also could examine the consequences and antecedents of latent, as opposed to measured, satisfaction. In particular, the estimation of the true association between latent satisfaction and customer lifetime value could contribute to the debate about the value of improving customer satisfaction (Bolton 1998; Kamakura et al. 2002). In general, we believe that any research that uses a survey that goes beyond description and examines the association between constructs can benefit from our method for testing for self-generated validity and other sources of measurement reactivity.

Appendix

We briefly describe the two-stage least squares estimator that Bollen (1996) introduces. The two-stage least squares estimator is consistent, allows nonnormal observed and latent variables, enables easy estimation of interaction effects, and has been used in many applications (for a review, see Schumacker and Marcoulides 1998).

The structural equation model represented in Figure 1 consists of a latent variable model (described in the text) and a measurement model, which can be expressed as follows:

$$\text{(A1)} \qquad MI = LI + \delta_{MI}, \text{ and}$$

$$RI = \alpha_{RI} + \lambda_{RI}(LI) + \delta_{RI},$$

where MI is measured intent as provided by the survey, LI is a latent variable that measures intent just before the time

[3]We thank an anonymous reviewer for noting that our method would also work in this circumstance.

[4]We thank the editor for noting the relationship between our work and previous research on calibration.

of the survey without error, RI is another reflective indicator of prior latent intent (the model can easily accommodate more indicators), α_{RI} is an intercept, and δ_{MI} and δ_{RI} are the two disturbance variables. Equation A1 shows that MI is set to have the same metric and origin as LI (by setting the intercept to zero and the factor loading to one) to provide a scale for the latent variable.

It is likely that some of the other indicators of prior latent intent are formative. We can express the relationship between a formative indicator (FI) and LI as follows:

$$\text{(A2)} \qquad LI = \alpha_{FI} + \lambda_{FI}(FI) + \zeta_{FI},$$

where ζ_{FI} is another disturbance term. We can extend Equation A2 to multiple formative indicators. Following the traditional assumptions of structural equation modeling, we assume that δ_{MI} and δ_{RI} are independent of LI and of each other and that LI, δ_{MI}, δ_{RI}, ζ_{FI}, and ε are each i.i.d. random variables. We also assume that δ_{MI}, δ_{RI}, ζ_{FI}, and FI are independent of S (a binary variable that measures whether consumers were surveyed).

Equation A1 shows that LI = MI – δ_{MI}. Substituting LI with MI – δ_{MI}, we obtain the following:

$$\text{(A3)} \qquad B = \alpha + \beta_1(MI) + \beta_2(S) + \beta_3(MI)(S) + \mu,$$

where B is the future behavior of interest (purchase incidence, customer profitability), and μ is a composite disturbance:

$$\text{(A4)} \qquad \mu = \varepsilon - \beta_1(\delta_{MI}) - \beta_3(S)(\delta_{MI}).$$

Equation A3 shows that the original latent variable model in Equation 1 can be rewritten as a model with only observed variables and a disturbance term μ. Because of measurement error, MI and δ_{MI} are correlated, and thus MI is correlated with the composite disturbance term μ, which violates the assumptions of the OLS estimator. Therefore, we must replace MI with an instrument variable that is correlated with MI but not with δ_{MI}. Other indicators of prior latent intent, whether reflective or formative, can be used as instrument variables in a two-stage procedure because they are not correlated with δ_{MI} and because, as other measures of latent intent, they are correlated with LI.

In the first stage, we regress MI on n other reflective indicators of LI (RI_1, RI_2, …, RI_n) and on the m other formative indicators of LI (FI_1, FI_2, …, FI_m), using data from the survey sample. In the second stage, we replace MI with its predicted value ($\hat{MI}$) in both samples. We then obtain the following equation:

$$\text{(A5)} \qquad B = \alpha_1 + \beta_1(\hat{MI}) + \beta_2(S) + \beta_3(\hat{MI})(S) + \mu.$$

Because $\hat{MI}$ is a linear combination of instrument variables (RI_1, RI_2, …, RI_n; FI_1, FI_2, …, FI_m), it is uncorrelated with δ_{MI} and ε and thus with μ. Therefore, Equation A5 can be estimated through a regular OLS regression to obtain β_1, β_2, and β_3, the parameters of interest.

REFERENCES

Armstrong, J. Scott and Fred Collopy (1992), "Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons," *International Journal of Forecasting*, 8 (1), 69–80.

Baumgartner, Hans and Jan-Benedict E.M. Steenkamp (1992), "Response Styles in Marketing Research: A Cross-National Investigation," *Journal of Marketing Research*, 38 (May), 143–56.

Bemmaor, Albert C. (1995), "Predicting Behavior from Intention-to-Buy Measures: The Parametric Case," *Journal of Marketing Research*, 32 (May), 176–91.

Bollen, Kenneth A. (1996), "An Alternative Two-Stage Least Squares (2SLS) Estimator for Latent Variable Equations," *Psychometrika*, 61 (1), 109–121.

——— and Pamela Paxton (1998), "Two-Stage Least Squares Estimation of Interaction Effects," in *Interaction and Nonlinear Effects in Structural Equation Modeling*, Randall E. Schumacker and George A. Marcoulides, eds. London: Lawrence Erlbaum Associates.

Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction," *Marketing Science*, 17 (1), 45–65.

Chandon, Pierre, Vicki G. Morwitz, and Werner J. Reinartz (2004), "The Short- and Long-Term Effects of Measuring Intent to Repurchase," *Journal of Consumer Research*, 31 (3), 566–72.

Dholakia, Paul M. and Vicki G. Morwitz (2002a), "How Surveys Influence Customers," *Harvard Business Review*, 80 (5), 18–19.

——— and ——— (2002b), "The Scope and Persistence of Mere-Measurement Effects: Evidence from a Field Study of Customer Satisfaction Measurement," *Journal of Consumer Research*, 29 (2), 159–67.

Feldman, Jack M. and John G. Lynch Jr. (1988), "Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior," *Journal of Applied Psychology*, 73 (3), 421–35.

Fishbein, Martin and Icek Ajzen (1975), *Beliefs, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Fitzsimons, Gavan J. and Vicki G. Morwitz (1996), "The Effect of Measuring Intent on Brand-Level Purchase Behavior," *Journal of Consumer Research*, 23 (1), 1–11.

Greenleaf, Eric A. (1992), "Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles," *Journal of Marketing Research*, 29 (May), 176–88.

Hsiao, Cheng, Bao Hong Sun, and Vicki G. Morwitz (2002), "The Role of Stated Intentions in New Product Purchase Forecasting," in *Advances in Econometrics*, Vol.16, Thomas B. Fomby and R. Carter Hill, eds. Oxford: Elsevier Science, 11–28.

Irwin, Julie R. (1999), "Introduction to the Special Issue on Ethical Trade-Offs in Consumer Decision Making," *Journal of Consumer Psychology*, 8 (3), 211–13.

——— and Gary H. McClelland (2001), "Misleading Heuristics and Moderated Multiple Regression Models," *Journal of Marketing Research*, 38 (February), 100–109.

Jamieson, Linda F. and Frank M. Bass (1989), "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods," *Journal of Marketing Research*, 26 (August), 336–45.

Juster, F. Thomas (1966), "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design," *Journal of the American Statistical Association*, 61 (September), 658–96.

Kalwani, Manohar U. and Alvin J. Silk (1982), "On the Reliability and Predictive Validity of Purchase Intention Measures," *Marketing Science*, 1 (3), 243–86.

Kamakura, Wagner A., Vikas Mittal, Fernando De Rosa, and Jose Afonso Mazzon (2002), "Assessing the Service-Profit Chain," *Marketing Science*, 21 (3), 294–317.

Manski, Charles F. (1990), "The Use of Intentions Data to Predict Behavior: A Best-Case Analysis," *Journal of the American Statistical Association*, 85 (412), 934–40.

Mittal, Vikas and Wagner A. Kamakura (2001), "Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics," *Journal of Marketing Research*, 38 (February), 131–42.

Morrison, Donald G. (1979), "Purchase Intentions and Purchasing Behavior," *Journal of Marketing*, 43 (April), 65–74.

Morwitz, Vicki G. (2001), "Methods for Forecasting from Intentions Data," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. Scott Armstrong, ed. Norwell, MA: Kluwer Academic Publishers, 34–56.

——— and Gavan J. Fitzsimons (2004), "The Mere Measurement Effect: Why Does Measuring Intentions Change Actual Behavior?" *Journal of Consumer Psychology*, 14 (1–2), 64–74.

———, Eric Johnson, and David Schmittlein (1993), "Does Measuring Intent Change Behavior?" *Journal of Consumer Research*, 20 (1), 46–61.

Reichheld, Frederick F. (2003), "The One Number You Need to Grow," *Harvard Business Review*, 81 (12), 46–57.

Reinartz, Werner J. and V. Kumar (2000), "On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, 64 (October), 17–35.

Rossi, Peter E., Zvi Gilula, and Greg M. Allenby (2001), "Overcoming Scale Usage Heterogeneity: A Bayesian Hierarchical Approach," *Journal of the American Statistical Association*, 96 (453), 20–31.

Schumacker, Randall E. and George A. Marcoulides (1998), *Interaction and Nonlinear Effects in Structural Equation Modeling*. London: Lawrence Erlbaum Associates.

Sheppard, Blair H., Jon Hartwick, and Paul R. Warshaw (1988), "The Theory of Reasoned Action: A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research," *Journal of Consumer Research*, 15 (3), 325–43.

Sherman, Steven J. (1980), "On the Self-Erasing Nature of Errors of Prediction," *Journal of Personality and Social Psychology*, 39 (August), 211–21.

Simmons, Carolyn J., Barbara A. Bickart, and John G. Lynch Jr. (1993), "Capturing and Creating Public Opinion in Survey Research," *Journal of Consumer Research*, 20 (2), 316–29.

Spangenberg, Eric R. (1997), "Increasing Health Club Attendance Through Self-Prophecy," *Marketing Letters*, 8 (1), 23–32.

——— and Anthony G. Greenwald (1999), "Social Influence by Requesting Self-Prophecy," *Journal of Consumer Psychology*, 8 (1), 61–89.

Sprott, David E., Eric R. Spangenberg, Bianca Grohmann, and Ronn J. Smith (2003), "Mass-Communicated Prediction Requests: Practical Application and a Cognitive Dissonance Explanation for Self-Prophecy," *Journal of Marketing*, 67 (July), 47–62.

Sudman, Seymour and Brian Wansink (2002), *Consumer Panels*, 2d ed. Chicago: American Marketing Association.

Wansink, Brian and Michael L. Ray (2000), "Estimating an Advertisement's Impact on One's Consumption of a Brand," *Journal of Advertising Research*, 40 (6), 106–113.