



Published in final edited form as:

*Soc Psychol Q.* 2016 December ; 79(4): 333–354. doi:10.1177/0190272516628298.

## Lies, Damned Lies, and Survey Self-Reports? Identity as a Cause of Measurement Bias

Philip S. Brenner<sup>1</sup> and John DeLamater<sup>2</sup>

<sup>1</sup>University of Massachusetts Boston, Boston, MA, USA

<sup>2</sup>University of Wisconsin–Madison, Madison, WI, USA

### Abstract

Explanations of error in survey self-reports have focused on social desirability: that respondents answer questions about normative behavior to appear prosocial to interviewers. However, this paradigm fails to explain why bias occurs even in self-administered modes like mail and web surveys. We offer an alternative explanation rooted in identity theory that focuses on measurement directiveness as a cause of bias. After completing questions about physical exercise on a web survey, respondents completed a text message–based reporting procedure, sending updates on their major activities for five days. Random assignment was then made to one of two conditions: instructions mentioned the focus of the study, physical exercise, or not. Survey responses, text updates, and records from recreation facilities were compared. Direct measures generated bias—overreporting in survey measures and reactivity in the directive text condition—but the nondirective text condition generated unbiased measures. Findings are discussed in terms of identity.

### Keywords

identity; measurement; survey research; time diaries; physical exercise

---

Survey estimates of normative behavior—like voting, exercising, and church attendance—often include substantial measurement error as respondents report higher rates of these behaviors than is warranted (Bernstein, Chadha, and Montjoy 2001; Hadaway, Marler, and Chaves 1998; Shephard 2003). Conventional wisdom suggests that this error, called social desirability bias, is generated by a respondent’s need to appear prosocial. Because these behaviors are valued and widely seen as good—for the individual, his or her community, or society—they are claimed on surveys even when the respondent’s behavior does not support such claims.

This common understanding suggests that reports of normative behavior will not be made equally on all types of surveys. Rather, respondents are expected to report a higher frequency on interviewer-administered surveys than on the self-administered questionnaires used in mail and web surveys. This understanding is borrowed from a similar but distinct

phenomenon, the underreporting of counternormative behavior. Lower rates of counternormative behaviors, like drug use and abortion, are reported to an interviewer than on self-administered surveys (Tourangeau and Yan 2007). Underreporting of these behaviors in a survey interview is commonly attributed to impression management—it is relatively easier to admit a “bad” truth on a paper or computerized questionnaire than to a human interviewer.

But overreporting of normative behavior is a distinct phenomenon, not simply the obverse of underreporting of counternormative behavior. Unlike the embarrassment of admitting to an interviewer a drunk driving arrest or illegal substance use, the desire to claim normative behavior is not necessarily assuaged by shifting from an interviewer-administered to a self-administered mode. Evidence suggesting that interviewer-administered modes generate more response bias than do self-administered modes for questions about normative behavior is limited. While Stocké (2007) and Holbrook and Krosnick (2010) found higher rates of overreporting of voting in an interviewer-administered than a self-administered mode, other studies suggest that response validity in measures of normative behavior is not necessarily improved by self-administration. Kreuter, Presser, and Tourangeau (2008) found differences in the hypothesized direction in reported rates of four undesirable characteristics (including getting an unsatisfactory grade and receiving an academic warning) between interviewer- and self-administered modes, but similar differences failed to emerge for a set of five normative behaviors (including receiving academic honors and donating money to the university). Recent research by the Pew Research Center (2015a) found little difference in the reported frequency of church attendance from respondents assigned randomly to a telephone interview or a web survey.

Moreover, evidence suggests that substantial bias emerges in self-administered modes thought to be less susceptible to impression management. Comparisons of survey reports to criterion measures (administrative records and diary-like measures) found self-reported rates of exercise and church attendance that were double the actual frequency of these behaviors (Brenner and DeLamater 2014, forthcoming). Finding high rates of over-reporting of normative behavior even in self-administered modes calls into question the conventional wisdom that this phenomenon is simply the mirror image of the underreporting of counternormative behavior rooted in impression management. In summary, the evidence for conventional explanations of the overreporting of normative behavior fails to fully explain the phenomenon. But if impression management fails to explain this survey error, what can?

In this article, we present an alternative explanation drawing on fundamental concepts in social psychology. We focus on bias in self-reports of physical exercise given evidence for the phenomenon from epidemiology and social science. Epidemiologists have compared survey self-reports with physical measures that provide potential evidence of physical activity (Bassett 2000; Patterson 2000).<sup>1</sup> Social science methods, however, more directly estimate bias in survey self-reports. Using flexible interviewing techniques, Rzewnicki,

---

<sup>1</sup>These include percentage body fat (Ainsworth, Jacobs, and Leon 1992); exercise performance data recorded by treadmills, accelerometers and actigraphs (Adams et al. 2005; Jacobs et al. 1992; Leenders et al. 2001; Matthews and Freedson 1995); estimation of metabolic rate using doubly labeled water (Adams et al. 2005); and estimation of energy expenditure from the caloric intake of a controlled feeding program (Albanes et al. 1990).

Auweele, and De Bourdeaudhuij (2003) reinterviewed a random subsample of 50 survey respondents about time spent walking and sitting in the past seven days; 23 of the 50 claimed less time walking than in the original interview. Using direct but unobtrusive observation, Klesges et al. (1990) found that respondents who reported high levels of habitual physical activity seven days prior to a survey showed the greatest divergence between their self-reported and actual levels of activity during a one-hour period during which their activity was surreptitiously observed. Chase and Godbey (1983) discovered high rates of overreporting when comparing survey reports of physical activity (tennis and swimming) with data from recreation facility sign-in sheets.

Accurate measurement of exercise behavior is important in both public health and social psychology since it is a major influence on health and wellness. Following the presentation of the model, we present data from an experiment designed to test key propositions. The experiment introduces the use of contemporary technology, cell phones and short message service (SMS) text messaging, for data collection and the use of a “gold standard” criterion to assess the validity of self-reports. Finally, we discuss both the methodological and substantive contributions of our work.

## Identity Theory and Measurement

Identity theory is derived from a symbolic interactionist approach to social behavior that gives primacy to structure over agency (Stryker 1980). The theory conceptualizes the social relationships of the individual in terms of the roles he or she fills in interaction with others. In its original formulation, identity theory posits identities as, “‘parts’ of self, internalized positional designations [that] exist insofar as the person is a participant in structured role relationships” (Stryker 1980:60). An individual is comprised of a multiplicity of identities, the number of which is determined by the roles one fills in interaction with others (James [1890] 1950). An individual may have as many identities as he or she has roles, each contributing to the composition of the self. This original conceptualization of the theory, now referred to as its structural version, is complemented by a perceptual control version (Stryker and Burke 2000), which focuses on the verification of identities and the consequences of the success or failure of the verification process (Burke 1991; Stets and Burke 2005). Moreover, in addition to role identities, identity theorists have integrated group and person identities into the theory (Burke and Stets 2009; Stets and Burke 2000; Stets and Serpe 2013).

Identity theory posits two concepts, prominence and salience, to understand the functioning of identities that may be helpful in understanding the role of identity in survey measurement. The *prominence* of an identity is the individual’s affective connection to it, reflecting an idealized or aspirational version of the self (McCall and Simmons 1978). Prominence is often used synonymously with the closely related concepts of psychological centrality (Rosenberg 1979) and importance (Ervin and Stryker 2001), each reflecting the subjective valuation of an identity to the self and its relative ranking in the self-concept. Thus, identities are arranged in a prominence hierarchy, a high placement reflecting an identity’s high subjective value to the individual’s ideal self-concept—the person he or she wants to be (Brenner, Serpe, and Stryker 2014; McCall and Simmons 1978). While the individual is by

definition aware of the value placed on an identity, he or she does not necessarily know his or her likelihood of performing that identity (Stryker and Serpe 1994).

The second concept, *salience*, is the probability of an identity being enacted or the propensity to define a situation as one relevant for identity enactment (Stryker 1980; Stryker and Serpe 1994). Measures of salience have typically asked respondents to report on likely behavior in hypothetical situations (Brenner et al. 2014; Serpe 1987), although recent work has used self-reported frequency of identity-related behavior (Brenner 2011a, 2012, 2014). Identities are similarly arranged hierarchically by salience, a high placement signifying its high likelihood of enactment. As these definitions suggest, identity theory locates prominence causally prior to salience (Ervin and Stryker 2001). The prominence of the identity informs its likelihood of being enacted; highly prominent identities are more likely to be enacted (Brenner et al. 2014).

This conceptualization of identity has direct implications for survey measurement. While levels of identity prominence and salience typically agree, they need not. Congruence between prominence and salience may be an artifact of measurement (Stryker and Serpe 1994). In addition to promoting identity enactment in conventional settings (e.g., working out at the gym), prominence can also affect the measurement of salience in a way that encourages bias. The respondent who values physical exercise and sees himself or herself as the “kind of person” who is physically active may not enact the exercise identity at a rate consistent with the identity given the costs of its enactment (e.g., the time needed for a workout or the monetary cost of a gym membership). However, the individual may interpret the survey interview as a low cost opportunity to enact the identity. According to Burke (1980:28), “the problem with most measurement situations is that without the normal situational constraints it becomes very easy for a respondent to give us that idealized identity picture which may only seldom be realized in normal interactional situations.”

Thus, conventional direct survey questions can prompt the respondent to reflect not only on the actual self—the self realized in daily interactional situations—but also on the person he or she wishes to be (ideal self) or thinks he or she ought (ought self) to be (Higgins 1987). These three domains of the self—actual, ought, and ideal—can influence survey response. The ideal self encompasses behavior relevant to identities to which the individual aspires; the person one wishes to be (Large and Marcussen 2000; Marcussen and Large 2003, 2006). The ought self domain reflects the normative identities and their concomitant behavior valued by the society, community, and groups of which the individual is a member. Thus, the ought self is the person one believes he or she ought to be given these internalized norms. Potentially in contrast to the ideal and ought selves, the actual self domain includes the identities the individual regularly or typically enacts. The actual self is situationally and structurally constrained in ways that can prevent performance of behavior that verifies prominent identities. These constraints, including time, money, physical ability, and so on, can limit the extent to which the actual self reflects the ought self or ideal self.

Underreporting of counternormative behaviors is primarily motivated by the ought self. The respondent reports lower rates of these behaviors to an interviewer than warranted in order to appear socially desirable (Tourangeau and Yan 2007). Conversely, overreporting of

normative behavior is primarily motivated by the ideal self, although not necessarily deliberately (Brenner 2011a, 2012; Hadaway et al. 1998). Rather than being motivated solely by self-presentational concerns, the respondent pragmatically reinterprets the question (Clark and Schober 1992; Schwarz 1996; Sudman, Bradburn, and Schwarz 1996) to be one about identity rather than behavior (Hadaway et al. 1998), a process influenced by a desire for consistency between the ideal self and the actual self. This pragmatic interpretation of the survey question encourages the respondent to answer in a way that affirms strongly valued identities.

This process is related to but distinct from self-verification as it is typically conceptualized. Like self-verification, the identity process as applied to survey measurement allows the respondent to “create—both in the actual social environments and in their own minds—a social reality that verifies and confirms their self-conceptions” (Swann 1983:33). However, the verification process focuses on seeking and attending to others’ confirmatory feedback on prominent identities (Burke 1991; Swann, Pelham, and Krull 1989). Thus, the individual has a strong desire to ensure the consistency of others’ views, given their feedback, with his or her own. But self-administered surveys do not provide such feedback. The survey data collection process, especially in its self-administered formulation, is a one-way “exchange” of information, from the respondent to the data collector (i.e., an interviewer, webpage, or paper questionnaire), leaving no room for verification—except, and importantly, for that starting and ending in the self. Such self-verification completed by the individual is better understood through the lens of identity theory.

Unlike conventional survey questions, chronological data collection procedures, like time diaries, sidestep the bias contributed by prominent identities (Bolger, Davis, and Rafaeli 2003). By eschewing direct questions about specific behaviors of interest (Robinson 1985, 1999; Stinson 1999), chronological data collection procedures avoid prompting identity-related self-reflection on the part of the respondent. Thus, they yield less biased, higher quality data on many normative behaviors (Bolger et al. 2003; Niemi 1993; Zuzanek and Smale 1999). Work by Presser and Stinson (1998) and Brenner (2011b) compared survey estimates of a normative behavior, church attendance, with those from time diary studies. Both studies find reduced reporting of normative behavior relative to conventional survey reports, arguably because they do not invite reactivity (behavior altered as a result of observation) by focusing the respondent’s attention to any particular behavior.

Brenner and DeLamater (2013, forthcoming) extended this approach to examine the overreporting of physical exercise. They tested a novel chronological data collection procedure using SMS text messaging and compared it to a conventional survey question administered via web survey. Both data collection methods measured frequency of physical exercise in two subsamples of undergraduates during a one-week reference period. Respondents in the text-message condition were asked to report for the five days following enrollment each change in activity. Respondents completing the web survey were asked on how many of the seven days preceding enrollment they exercised overall and at campus recreation facilities. Validation with records from university recreation facility admissions (i.e., ID card scans upon entrance) uncovered high rates of overreporting on the survey; half of the respondents overreported their frequency of exercise over the past week (Brenner and

DeLamater forthcoming).<sup>2</sup> In contrast, the chronological reporting procedure using SMS resulted in high-quality, unbiased reports of exercise frequency.

These findings suggest that direct measurement can be directive and a cause of error. Direct, retrospective measures of normative behavior, like those on conventional surveys, are hypothesized to generate bias as respondents reflect on the ideal self in answering these questions. Indirect, prospective chronological measurement procedures avoid bias by not focusing the respondent's attention on any particular behavior. However, this hypothesis has not been tested experimentally by manipulating the directiveness of the chronological measurement procedure.

## HYPOTHESES

This study importantly extends previous research and addresses key limitations. Unlike prior investigations (in which respondents reported exercise using only one mode), respondents in the current study answer *both* conventional survey questions about physical exercise and participate in a chronological measurement procedure. This improvement in the study design allows us to predict bias within (for each respondent) rather than between (the marginal difference between conditions). These measures are compared to each other and to a reverse record check from recreation facilities for activity reported at these sites to estimate measurement bias. A statistically and substantively significant overreport is expected when comparing survey self-reports of exercise to criterion measures from facility records (for on-campus exercise) and chronological reports.

*Hypothesis 1:* Respondents' survey reports of exercise will be higher than either their chronological reports of exercise or their admission records from campus recreation facilities.

Behavior is encouraged by identity prominence; in short, we tend to perform identities that we value (Brenner et al. 2014). Thus, respondents with high exercise identity prominence are strongly motivated to perform that identity. But the gymnasium (or the basketball or tennis court, hiking trail, rock wall, etc.) is not the only site for the performance of the exercise identity. Understanding survey measurement as an opportunity for identity performance helps to explain the occurrence of overreporting even on self-administered surveys. Brenner (forthcoming) demonstrates that costs can situationally constrain identity performance. The respondent who values his or her identity as a physically active individual may fail to perform it given costs like time (i.e., an hour at the gym) and money (i.e., an expensive gym membership). But when presented an inexpensive opportunity to perform the identity by simply answering a survey question in the affirmative, he or she may take this opportunity to do so. Thus, if identity prominence promotes the performance of the exercise identity in both situations—encouraging actual physical activity and the survey reports of it—the difference in identity prominence between respondents who validly report exercise and those who overreport their exercise will be minimal. The obverse is a comparison of non-exercisers. If prominence motivates the claim of (unwarranted) behavior on a survey, a

---

<sup>2</sup>This brief, well-defined reference period addresses the concern of memory failure and its resulting reporting error addressed by Belli et al. (1999, 2001) and others.

difference should emerge between respondents with high prominence (who will be more likely to overreport) and low prominence (who will be less likely to overreport).

*Hypothesis 2:* Exercise identity prominence will be a strong predictor of overreporting among nonexercisers but not among self-reported exercisers.

Measurement directiveness triggers identity-related bias, but nondirective measurement, like that used in chronological measurement procedures, sidesteps the bias inherent in directive survey questions on normative behavior. If chronological measurement is altered to be directive and retrospective, it too likely generates overreporting of normative behavior. Such a measure asks specific questions about particular normative behaviors performed during the previous day or week, prompting identity-related self-reflection and resulting in overreporting on par with conventional survey questions. But if the directive chronological measurement procedure were in situ, respondents would have the opportunity to change their behavior to match their ideal or ought self. This error is called *measurement reactivity*.<sup>3</sup> Thus, random assignment to either (1) a standard nondirective chronological or (2) an experimental directive chronological measurement condition is used to estimate the effect of directiveness of measurement—a test of an important assumption of previous work and extension of the extant research.

Respondents receiving directive chronological measurement instructions are hypothesized to alter their behavior as a result of the priming effect of directive measurement that encourages normative behavior, yielding reactivity. This effect may potentially arise in interaction with prominence as those with highly prominent exercise identities may be additionally motivated to change their behavior if randomly assigned to the directive condition.

*Hypothesis 3:* Respondents receiving a directive chronological measure will be more likely to exercise during the reference period than those receiving a nondirective measure.

## METHODS

In February 2014,<sup>4</sup> a random sample of 1,200 undergraduates at a large, public Midwestern university received an email invitation to participate in a two part study—a brief web survey followed by a five-day texting component—and contained a link to the first assessment. Respondents were promised a \$50 incentive upon completion of the study in its entirety. The web survey, titled “[University Name] Student Daily Life Study,” presented about 20 questions in the context of a survey of daily life in order to mask the true focus of the study. Questions asked about students’ use of campus libraries, the student union, and other campus facilities; participation in activities and attendance at events on campus; and patronage of establishments in the city. In this context, questions were included to measure the focal activity, physical exercise, both on and off campus. A battery of items was also included to measure the prominence of a set of key identities (e.g., student, family, work/

---

<sup>3</sup>This is an error because it causes the observed value to differ from the true value, defined as the frequency of behavior that would have occurred without the directive measurement procedure.

<sup>4</sup>The field period was chosen to maximize the share of exercise taking place in indoor facilities to allow for validation. Daily low temperatures ranged from  $-11^{\circ}\text{F}$  to  $+10^{\circ}\text{F}$  and highs from

occupation, religion) including the sports/exercise identity. Three hundred ninety students completed the survey (33 percent; AAPOR Response Rate #5 [American Association for Public Opinion Research 2015]).

Survey respondents were then randomly assigned to one of two conditions for the texting component of the study. (See Figure 1 for an example of the study sequence.) Respondents in both conditions received an email thanking them for completing the survey and instructing them how to complete the next phase of the study in which they would send text messages to update study staff on their major activities for the subsequent five days. These reports were described as similar to “tweets” (Twitter updates) or Facebook status updates. Respondents were asked to text an update when they began an activity, identifying the activity and its location. A two-page training document included a page of simple instructions with examples and a page with a “frequently asked questions” list. Examples shared with study participants included common activities (e.g., going to class at Science Hall, studying at the library, getting coffee at Starbucks, or working out at the [name of campus recreation building]). Respondents received a reminder email at the beginning of each texting day.

The two conditions differed in only one way. Condition 1 respondents were not told the true focus of the study, but Condition 2 respondents’ training letter included the following additional sentence: “While we want to hear about all your daily activities, we are specifically interested in your physical fitness activities—playing sports, working out, and being physically active.” This sentence was set in bold face in the first paragraph on the first page of the training document. A total of 327 students completed the texting phase (161 and 166 in each condition, respectively), yielding a compliance rate of 84 percent of the initial survey respondents.

Finally, at the conclusion of the study, participants were asked for permission to access their records of recreational sports facilities usage. A total of 285 students (143 and 142 in each condition, respectively) allowed access (87 percent compliance rate among texters), yielding a final response rate of 24 percent.<sup>5</sup> Respondents sent a total of over 6,100 updates, averaging over 20 per student and over 4 per day.

### Data and measures

Six measures of physical activity were collected using three methods. The initial web survey produced two measures. Respondents were first asked, “In the past seven days, how many days have you worked out or exercised?” This question was followed by a second question that asked, “Of the days you worked out in the past week, how many did you use [university recreational sports department name] facilities, like [names of facilities]?” These two questions measured exercise (1) overall and (2) specifically at campus facilities.

Two measures of physical activity were produced from respondents’ text messages. Texts were coded for exercise (3) overall and (4) specifically at campus facilities. The survey,

---

<sup>5</sup>Nine ineligible cases were omitted. Seven did not own text-capable cell phones, and two were current or former undergraduate research assistants for the principal investigators.



which was a retrospective measure, and text message reports, which were in situ, measure activity in consecutive, not matching, reference periods (see Figure 1).

The two final measures of exercise are criterion measures from a reverse record check of the facility admission database. Students entering the facilities present their student ID card to be scanned. Each record contains the time of admission and the student's name and ID number. Record data were accessed for each participant for comparison with both the (5) survey report and (6) text message reports, reflecting their differing reference periods (see Figure 1).<sup>6</sup>

An indicator of response validity is then computed as the difference between the survey measure and the criterion, from facility records or from the text report. The record check can only be used as the criterion for reports of exercise at campus facilities. The text report is used as the criterion for survey reports of exercise without consideration of location. A positive difference defines the respondent as an *overreporter*. Overreporters can be further defined as *exercising* and *nonexercising overreporters* if their survey report is greater or equal to zero, respectively. Equivalence is defined as a valid response either as a *validated exerciser* or an *admitted nonexerciser* (see Figure 2).

Exercise identity prominence was measured in the survey using the question "Each of us is involved in different roles and activities. How important to you is exercise, working out, or playing sports?" The prominence of the other identities common in this population was also assessed (e.g., being a student, a member of one's family, a member of a student group). Respondents were presented an 11-point response scale from 0 = not at all important to 10 = extremely important for each identity. We use a single-item measure of exercise identity prominence in the context of other identities to prevent identifying the main focus of the study to the respondent. A similar one-item measure was used by Callero (1985) in research on another prosocial identity (i.e., blood donor.) He found that the rank given that identity is highly correlated with a five-item measure of the prominence of blood donation in the person's life (e.g., "I would feel a loss if I were forced to give up donating blood.").

Finally, condition is a dichotomous indicator noting the respondent's random assignment into the treatment (including the directive statement) or control (nondirective) group.

### Analysis Plan

The number of days of exercise reported on the survey is compared to facility records and text reports using the Wilcoxon signed-rank test, a nonparametric analog to the paired *t*-test used because the dependent variable is ordinal and non-normally distributed. Identity prominence is compared between groups of respondents using the Mann-Whitney-Wilcoxon rank-sum test, a nonparametric analog to Student's *t*. The hypothesis of equality between (1) the text reports and record data and (2) the identity prominence between overreporters and other self-reported exercisers is tested using the two one-sided tests (TOST) approach

---

<sup>6</sup>Pretest ethnographic research demonstrated the quality of record data. A research assistant visited each campus recreation facility, varying observations by day and time. Persons entering the facility were observed to ensure that an ID card was scanned and were then followed long enough to ensure that their use of the facility was exercise or sport. All those admitted to the facility presented an ID card and all proceeded to exercise, often after a stop in the locker room.

(Barker et al. 2002; Wellek 2010). A measure of effect size,  $r$ , computed as the ratio of the test statistic  $z$  and the square root of the sample size, is used to understand the substantive significance of any differences from these comparisons. Finally, logistic regression models predict texted reports of exercise using (1) condition (directive or nondirective texting instructions), (2) the prominence of the respondent's exercise identity, and (3) their interaction.

## RESULTS

### Validating the Text Reports

The first step of this analysis is comparing the two criterion measures: the texted reports of exercise and the recreation facility admission records. Establishing the text report as a valid measure of on-campus exercise (by comparing to facility admission records) adds credence to its use as a criterion of exercise outside these facilities—locations where admission records are not available (i.e., private or commercial gyms) or do not exist (i.e., outdoor activities like jogging and cross-country skiing).

Respondents reported in their text messages about one day of exercise at campus facilities over the five-day field period ( $\mu = .92$ ,  $\sigma = 1.13$ ), very similar to that from admission records from the campus facilities ( $\mu = .85$ ,  $\sigma = 1.18$ ). The text report and admission records fail to differ statistically ( $z = 1.7$ , *ns*; see Figure 3) or substantively ( $r = .10$ ). But the lack of a significant difference does not necessarily indicate their equivalence. Therefore, equivalence between these two measures is tested applying the TOST procedure, using a small effect size ( $r = .20$ ; approximately a quarter-of-a-day difference) as the equivalence margin. The hypothesis of equivalence is supported ( $z_1 = 1.66$ ,  $p < .05$ ;  $z_2 = 5.10$ ,  $p < .001$ ). The comparability of these distributions suggests that the text reports adequately represent the true value of exercise where validation is possible. In the following analyses, text reports will be used as a criterion measure in conjunction with records. Importantly, text reports are used in lieu of facility records for comparison with reports of exercise outside campus facilities when record data do not exist.

### Comparison of Reported and Actual Exercise

Comparison of the three data collection procedures demonstrates the effect of directiveness of measurement. As hypothesized, a much higher rate of exercise at campus facilities is reported on the survey ( $\mu = 1.74$ ,  $\sigma = 1.83$ ) than can be validated in records ( $\mu = 1.08$ ,  $\sigma = 1.57$ ), the difference reaching conventional levels of statistical significance ( $z = 7.8$ ,  $p < .001$ ; see Figure 3). Similarly, survey data show a higher rate of reported exercise when compared with text message data. Respondents texted about one day of exercise at campus facilities over the five-day field period ( $\mu = .92$ ,  $\sigma = 1.13$ ), much lower than the survey estimate of exercise at campus facilities ( $z = 8.3$ ,  $p < .001$ ).

Survey reports of overall exercise including both on- and off-campus exercise (including off-campus facilities and out-of-doors) can also be compared to text reports. Respondents texted approximately one day of exercise overall during the five-day field period ( $\mu = 1.14$ ,  $\sigma =$

1.20), much lower than the survey estimate of nearly three days ( $\mu = 2.78$ ,  $\sigma = 1.90$ ;  $z = 12.6$ ,  $p < .001$ ).

In summary, these comparisons suggest that respondents report a great deal more exercise on the survey than either facility records or chronological reports warrant. In each of these comparisons, the effect size of the difference between the survey and the criterion is medium to large ( $r = .46-.75$ ). These findings confirm Hypothesis 1.

### Predicting Overreporting

In the following analyses, identity prominence is compared across the categories of respondents (i.e., overreporters, admitted nonexercisers, and validated exercisers) to examine its association with behavior and response validity.<sup>7</sup> As hypothesized, prominence distinguishes between the two types of nonexercisers. Nonexercising overreporters report higher exercise identity prominence than do admitted nonexercisers ( $z = 3.69$ ,  $p < .001$ ;  $r = .29$ ; see Figure 4). Thus, prominence is strongly associated with unwarranted claims of exercise.

Additional comparisons demonstrate the similarity between overreporters and validated exercisers. No significant or substantive difference in identity prominence emerges when comparing overreporters to other respondents who validly reported their frequency of exercise ( $z = 1.11$ ,  $ns$ ;  $r = .08$ ), including overreporters who exercised at least once during the reference period but overreported its frequency ( $z = .78$ ,  $ns$ ;  $r = .08$ ). Equivalence between these two measures is tested using a small effect size ( $r = .20$ ; approximately half a point of prominence) as the equivalence margin. The hypothesis of equivalence is supported for the comparison of verified exercisers with overreporters ( $z_1 = 4.5$ ,  $p < .001$ ;  $z_2 = 2.2$ ,  $p < .05$ ) and exercising overreporters ( $z_1 = 4.16$ ,  $p < .001$ ;  $z_2 = 2.60$ ,  $p < .01$ ). Thus, the prominence of respondents who exercise and those who overreport, whether they exercised or not, are equivalent. These findings confirm Hypothesis 2.

The final analyses focus on the effects of measurement directiveness, prominence, and their interaction. The role of the measurement condition is used as a predictor of exercise along with identity prominence to answer the following question: does the manipulation—explicitly mentioning the focus of the study in the document describing the text message procedure—change the behavior of the respondent? The interaction of condition and prominence is then tested to investigate the potential for differential effects of prominence in the two conditions.

Respondents in the directive condition (Condition 2) have nearly 20 times the odds of exercising compared to respondents in the nondirective condition (Condition 1; see Table 1). This model controls for exercise identity prominence, each additional point of which nearly doubles the respondent's odds of exercising. The model also includes a significant interaction between condition and prominence, flattening the probability curve for Condition 2 respondents relative to those in Condition 1.

<sup>7</sup>For simplicity of presentation, the criterion used here is from recreation facility records. Results are nearly identical if text-message data are used as the criterion instead.

To better understand this model, predicted probabilities at each prominence level are plotted separately for both conditions in Figure 5. The plotted probabilities for Condition 1 resemble a sigmoid function: respondents at low levels of exercise identity prominence have nearly a zero probability of exercise, and those at high levels of prominence approach a near certainty of exercise, with even odds of exercise falling just above the prominence midpoint. Comparably, the plotted Condition 2 predicted probabilities are nearly linear. At low prominence, Condition 2 respondents have nearly a 25 percent probability of exercise, increasing at 5 to 6 percentage points for each unit increase of prominence.

While our hypothesis (Hypothesis 3) that respondents in (directive) Condition 2 would increase their exercise relative to those in (nondirective) Condition 1 is supported, we are somewhat surprised by the pattern of difference between conditions. We anticipated that respondents with relatively higher prominence would alter their behavior and those with relatively lower prominence would not. However, we discovered the opposite. The differences between the predicted probabilities are large and significant for respondents with low and middling exercise (ratings from 0 to 6) identity prominence. Predicted probabilities of exercise do not differ statistically at high levels of prominence. In sum, it is respondents with low exercise identity prominence who are more likely to exercise when told they are being observed. Respondents who very strongly value their exercise identity are equally, and highly, likely to exercise in each of the two conditions.

## DISCUSSION

These findings highlight the difficulty of validly measuring normative behavior. Measurement errors skew survey estimates of physical exercise as well as reports of voting (Belli, Traugott, and Beckmann 2001; Bernstein et al. 2001) and recycling (Barker et al. 1994; Chao and Lam 2011; Corral-Verdugo 1997). Estimates of religious behaviors like prayer (Brenner 2014) and church attendance (Brenner 2011b; Hadaway et al. 1998; Presser and Stinson 1998) are also skewed in a similar fashion. These errors are not universal. Rather, when and where measurement biases emerge is determined by the norms of the society, community, and groups being measured (Schwarz, Oyserman, and Peytcheva 2010; Uskul, Oyserman, and Schwarz 2010). It is not simply a matter of universal human nature that we exaggerate normative behavior on surveys, but rather what is normative is socially constructed and these exaggerations reflect societal, community, and group norms internalized by the individual. For example, the groups whose members overreport exercise may not overreport religious behaviors, and vice versa, and other groups may overreport both or neither.

Survey respondents overreport their activity to bring into congruence the ideal self, which includes many such internalized norms, and the survey report of behavior; in essence, bringing prominence and salience into consonance. The equivalent ratings of prominence between the two groups of self-reported exercisers from the survey—validated exercisers and overreporters—demonstrate the role of prominence in survey measurement and mismeasurement. For validated exercisers, the value placed on the identity encourages the enactment of the exercise identity by engaging in behaviors like physical exercise and playing sports. For overreporters, prominence is strong enough to encourage one type of

identity enactment (survey report of exercise) but fails to result in the other (actual physical exercise). But why do these groups diverge in their identity performances if they are similar in their prominence?

The role of prominence in encouraging identity performance is crucial to recent research in identity theory. Prominence is causally prior to salience; the higher the prominence, the more likely identity performance (Brenner et al. 2014; Stryker 1980). That prominence is rated (1) higher for overreporters than for validly reporting nonexercisers but (2) equivalent for all self-reported exercisers whether their report is valid or not yields additional evidence of the causal order of prominence to salience. If the reverse were true—that salience precedes prominence—behavior would determine valuation. In this case, we would expect that nonexercising overreporters would rate their prominence on par with other nonexercisers rather than exercisers.

This prominence-behavior link assumes that identity performance is at a situationally acceptable cost. Burke (1980) implies that “normal interactional situations” can introduce costs that inflate the expense of identity performance. For the exercise identity, typical types of performance are costly in terms of time and effort. In comparison, the conventional survey question itself provides a relatively low cost (or even negative cost as participants received a generous incentive) opportunity to enact an identity by making behavioral claims. For performances of the exercise identity at a gym, it may be that overreporters and validated exercisers differ in the costs they encounter or that validated exercisers are somehow better at overcoming those costs.

But when respondents encounter directive in situ measurement, reactivity is the result rather than overreporting. Respondents change their behavior to be consistent with or exceed identity claims. Measurement directiveness interacted with prominence to increase behavior rather than just reports of behavior. While both prominence and the manipulation had strong main effects promoting exercise for all respondents, their interaction surprisingly altered the relationship, widening the difference between the two conditions at low levels of prominence and closing it at higher levels. Thus, respondents with low exercise identity prominence changed their behavior in a normative direction when told the focus of the study.

This raises a crucial question for identity theory and measurement—why would directive measurement encourage individuals to perform an identity of low prominence? Individuals internalize the norms of the society, community, or group into which they are socialized (Mead 1934). These norms may be internalized to the point of becoming a component of the individual’s self-schema, specifically an idealized identity of high prominence and part of the ideal self. Failing to live up to the ideal self has ramifications: a discrepancy between the ideal and actual selves results in depression as the individual acutely feels the failure to live up to his or her goals (Higgins 1987). From this perspective, the need to avoid discrepancies between the actual and ideal selves may motivate measurement error in survey responses as respondents with high exercise identity prominence are much more likely to overreport than are respondents with low prominence. Direct survey measurement using conventional survey questions about normative behavior primes the respondent to consider not only actual behavior but also the value he or she places on the focal identity. Thusly primed, the

respondent risks depression if the highly prominent idealized identity fails to be confirmed. Therefore, the respondent pragmatically interprets the behavioral question to be about his or her identity and responds to this interpretation rather than the semantic meaning of the question based in actual behavior (Hadaway et al. 1998). In this way, overreporting is founded in prominence and the ideal self.

Alternatively, individuals may internalize societal, community, or group norms as socially desirable even if they are not central to one's own self-schema. According to social identity theory, the norms of the groups to which one belongs have ramifications for behavior, especially when group membership is made salient (Hogg, Terry, and White 1995; White, Terry, and Hogg 1994). The current study, titled (“[University Name] Student Daily Life Study”) and sent on university letterhead, explicitly directed respondents' attention to their identities as university students in order to hide the true focus of the study. Their student identity is one they value strongly: 87 percent rated this identity an 8 or higher on a scale from 0 to 10. In line with social identity theory, they are likely to internalize the norms of their group (students at an elite public university)—including the value placed on athletics and fitness (Terry, Hogg, and White 1999). Thus, even if their exercise identity is not prominent, the respondent may still have internalized the normativeness of athletics and exercise given the importance placed on membership in a group whose members value exercise (Terry and Hogg 1996). Failing to live up to the ought self also has ramifications: a discrepancy between the ought and actual selves results in social anxiety as the individual acutely feels the failure to live up to that which is valued by the group (Higgins 1987). But unlike the survey question asking about recent behavior, respondents text reports of behavior as it occurs, allowing the respondent to alter his or her behavior as it is measured to bring the actual and ought selves into alignment to avoid a discrepancy and subsequent anxiety. Thus, even individuals for whom the identity is not prominent may perform it given in situ, directive measurement.

### **Applications and Future Directions**

Understanding the cause of overreporting as rooted in identity rather than impression management sheds light on a potential solution. Attempts to reduce overreporting of normative behaviors using context effects (Presser 1990) and experimental question wordings (Belli, Traugott, and Rosenstone 1994; DeBell and Figueroa 2011) have achieved only limited success. The few successful experiments reduce but do not eliminate bias (Belli et al. 1999; Duff et al. 2007). Applying the findings of the current study may help to further reduce overreporting. If respondents' need to claim the identity may be fulfilled by one or more questions, asking about aspirational or perceived normative levels of behavior before the “actual” behavior question is presented may potentially produce a better estimate of behavior.

However, given these findings, it may be that nondirective—indirect or unobtrusive—measurement procedures are essential for unbiased and nonreactive measurement of normative behavior. The SMS procedure was demonstrated to be a useful tool for social psychologists specifically and sociologists more generally. While useful for studies of many types of frequently and regularly performed normative behaviors, SMS may be readily

adapted to other types of behavior and to the study of emotion, mood, or other mental states (Brenner and DeLamater 2014). Thus, SMS presents an alternative to the experience sampling method (ESM) and ecological momentary assessment (Larson and Csikszentmihalyi 1983). ESM is used to measure activity, mood, and other physical and mental states during sampled moments spanning one or more days. A beeper or text message flags the respondent to record the predetermined focal behavior, activity, mood, and so on. The SMS procedure offers two potential benefits over ESM. First, it sidesteps the priming effect of direct questions about particular normative behaviors, avoiding overreporting and reactivity. Second, allowing the respondent to control the timing of the report, rather than sampling some small number of moments for measurement, permits more adequate measurement of brief activities. For example, an ESM study interested in occasions of received social support may fail to sample brief periods of time that include key social interactions in which support may be received (e.g., conversations over a quick breakfast before work, brief coffee breaks with coworkers). The SMS procedure, however, may be readily adapted to nondirectively measure these types of activities.

Moreover, the SMS procedure fits well into the lives of many modern Americans and others in many countries. Nearly two-thirds of American adults own a smartphone, including 85 percent of young adults 18 to 29 years of age and 79 percent of adults ages 30 to 49 years of age (Pew Research Center 2015b). The SMS procedure may be a more appropriate measurement procedure than conventional time diaries for many in these populations, especially highly active individuals who make extensive use of smartphones (e.g., young adults and college students). For populations like these, using an inherently mobile data collection procedure is a better fit with their lifestyles. Moreover, by spreading the reporting burden out over the field period, the burden of the SMS reporting procedure may feel reduced compared to recording activities in a diary all at once.

## CONCLUSION

This article focuses on the lives we live and “lies” we tell. It addresses the nature of a very particular “lie”—overreporting of normative behavior on surveys. Survey respondents can give inaccurate answers to questions, especially those about normative behaviors. However, these responses may reflect other truths well as respondents treat the survey question as opportunity to truthfully report their self-views.

Direct measures of normative behavior, whether interviewer- or self-administered, can yield highly biased estimates. In the present study and elsewhere, high rates of bias emerge in survey estimates of normative behaviors like voting, attending religious services, and engaging in physical exercise. That extensive bias occurs even in a self-administered mode challenges the conventional wisdom that links overreporting of normative behavior to an impression management paradigm and attributes it to social desirability. But if impression management fails to explain this survey error, what can?

An approach founded in identity theory provides a more complete explanation of this phenomenon. With or without the presence of an interviewer, direct survey questions about normative behavior are pragmatically interpreted to be about the respondent’s identity,

asking whether he or she is the “kind of person” who conforms to the norm (Hadaway et al. 1998). In other words, the question is transformed from an inquiry about “what I do” to ask about “who I am.” Importantly, this self-view may not be rooted in the actual self. Rather, it may be strongly reflective of the ideal self—the person the respondent aspires to be. This transformation of the question allows identity prominence to exert its influence on the question answering process. If the relevant identity is one of high prominence, it provides a strong motivation to claim the behavior to bring into consonance the survey report and the self-view. Thus, our results suggest that a fundamental characteristic of the individual is intricately involved in the process of responding to social psychological measures.

While we as sociologists and survey researchers would prefer veracity from our respondents, we can benefit from the “opportunities for understanding” that measurement errors provide (Schuman 1982). Arguably, survey respondents are not necessarily motivated to dissemble. Rather, they answer the survey question in a way that tells us about their self-view. Talk may be cheap, but it is not meaningless. We can still learn a great deal about culturally situated human behavior from the errors in survey reports. And by approaching these errors seriously as phenomena worthy of study, we can better understand what they mean, why they occur, and perhaps develop the methods—like the SMS reporting procedure presented here—to prevent them.

## Acknowledgments

The authors would like to thank the staff at the Center for Demography and Ecology and the Department of Recreational Sports at UW Madison, four anonymous reviewers, and the editors.

### FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the Joseph P. Healey Research Grant Program at the University of Massachusetts Boston and the Conway-Bascom Fund at the University of Wisconsin-Madison.

## References

- Adams, Swann Arp, Matthews, Charles E., Ebbeling, Cara B., Moore, Charity G., Cunningham, Joan E., Fulton, Jeanette, Hebert, James R. The Effect of Social Desirability and Social Approval on Self-Reports of Physical Activity. *American Journal of Epidemiology*. 2005; 161:389–98. [PubMed: 15692083]
- Ainsworth, Barbara E., Jacobs, David R., Leon, Arthur S. Validity and Reliability of Self-Reported Physical Activity Status: The Lipid Research Clinics Questionnaire. *Medicine and Science in Sports and Exercise*. 1992; 25:92–98.
- Albanes, Demetrius, Conway, Joan M., Taylor, Philip R., Moe, Paul W., Judd, Joseph. Validation and Comparison of Eight Physical Activity Questionnaires. *Epidemiology*. 1990; 1:65–71. [PubMed: 2081242]
- American Association for Public Opinion Research. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 8. AAPOR; 2015.
- Barker, Kathleen, Fong, Lynda, Grossman, Samara, Quin, Colin, Reid, Rachel. Comparison of Self-Reported Recycling Attitudes and Behaviors with Actual Behavior. *Psychological Reports*. 1994; 75:571–77.
- Barker, Lawrence E., Luman, Elizabeth T., McCauley, Mary M., Chu, Susan Y. Assessing Equivalence: An Alternative to the Use of Difference Tests for Measuring Disparities in Vaccination Coverage. *American Journal of Epidemiology*. 2002; 156:1056–61. [PubMed: 12446263]



- Bassett, David R. Validity and Reliability Issues in Objective Monitoring of Physical Activity. *Research Quarterly for Exercise and Sport*. 2000; 71:30–36. [PubMed: 25680010]
- Belli, Robert F., Traugott, Michael W., Beckmann, Matthew N. What Leads to Voting Overreports? Contrasts of Overreporters to Validated Voters and Admitted Nonvoters in the American National Election Studies. *Journal of Official Statistics*. 2001; 17:479–98.
- Belli, Robert F., Traugott, Santa, Rosenstone, Steven J. Reducing Overreporting of Voter Turnout: An Experiment Using a Source Monitoring Framework. 1994. Retrieved January 14, 2016 (<http://www.electionstudies.org/resources/papers/documents/nes010153.pdf>)
- Belli, Robert F., Traugott, Michael W., Young, Margaret, McGonagle, Katherine A. Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring. *Public Opinion Quarterly*. 1999; 63:90–108.
- Bernstein, Robert, Chadha, Anita, Montjoy, Robert. Overreporting Voting: Why It Happens and Why It Matters. *Public Opinion Quarterly*. 2001; 65:22–44. [PubMed: 11264053]
- Bolger, Niall, Davis, Angelina, Rafaeli, Eshkol. Diary Methods: Capturing Life as It Is Lived. *Annual Review of Psychology*. 2003; 54:579–616.
- Brenner, Philip S. Identity Importance and the Overreporting of Religious Service Attendance: Multiple Imputation of Religious Attendance Using American Time Use Study and the General Social Survey. *Journal for the Scientific Study of Religion*. 2011a; 50:103–15.
- Brenner, Philip S. Exceptional Behavior or Exceptional Identity? Overreporting of Church Attendance in the US. *Public Opinion Quarterly*. 2011b; 75:19–41.
- Brenner, Philip S. Identity as a Determinant of the Overreporting of Church Attendance in Canada. *Journal for the Scientific Study of Religion*. 2012; 51:377–85.
- Brenner, Philip S. Testing the Veracity of Self-Reported Religious Behavior in the Muslim World. *Social Forces*. 2014; 92:1009–37.
- Brenner, Philip S. Narratives of Error from Cognitive Interviews of Survey Questions about Normative Behavior. *Sociological Methods & Research*. Forthcoming.
- Brenner, Philip S., DeLamater, John D. Predictive Validity of Paradata on Reports of Physical Exercise: Evidence from a Time Use Study Using Text Messaging. *Electronic International Journal of Time Use Research*. 2013; 10:38–54.
- Brenner, Philip S., DeLamater, John D. Social Desirability Bias in Self-Reports of Physical Activity: Is an Exercise Identity the Culprit? *Social Indicators Research*. 2014; 117:489–504.
- Brenner, Philip S., DeLamater, John. Measurement Directiveness as a Cause of Response Bias: Evidence from Two Survey Experiments. *Sociological Methods & Research*. Forthcoming.
- Brenner, Philip S., Serpe, Richard T., Stryker, Sheldon. The Causal Ordering of Prominence and Salience in Identity Theory: An Empirical Examination. *Social Psychology Quarterly*. 2014; 77:231–52. [PubMed: 27284212]
- Burke, Peter J. The Self: Measurement Implications from a Symbolic Interactionist Perspective. *Social Psychology Quarterly*. 1980; 43:18–29.
- Burke, Peter J. Identity Processes and Stress. *Social Psychology Quarterly*. 1991; 56:836–49.
- Burke, Peter J., Stets, Jan E. *Identity Theory*. New York: Oxford University Press; 2009.
- Callero, Peter. Role-Identity Salience. *Social Psychology Quarterly*. 1985; 48:203–15.
- Chao, Yu-Long, Lam, San-Pui. Measuring Responsible Environmental Behavior: Self-Reported and Other-Reported Measures and Their Differences in Testing a Behavioral Mode. *Environment and Behavior*. 2011; 43:53–71.
- Chase, David R., Godbey, Geoffrey C. The Accuracy of Self-Reported Participation Rates. *Leisure Studies*. 1983; 2:231–35.
- Clark, Herbert H., Schober, Michael F. Asking Questions and Influencing Answers. In: Tanur, JM., editor. *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage; 1992. p. 15-48.
- Corral-Verdugo, Victor. Dual Realities of Conservation Behavior: Self-Reports vs Observations of Re-Use and Recycling Behavior. *Journal of Environmental Behavior*. 1997; 17:135–145.
- DeBell, Matthew, Figueroa, Lucila. Results of a Survey Experiment on Frequency Reporting: Religious Service Attendance from the 2010 ANES Panel Recontact Survey. Paper presented at the

- 66th Annual Conference of the American Association for Public Opinion Research; Phoenix, AZ. 2011.
- Duff, Brian, Hanmer, Michael J., Park, Won-Ho, White, Ismail K. Good Excuses: Understanding Who Votes with an Improved Turnout Question. *Public Opinion Quarterly*. 2007; 71:67–90.
- Ervin, Laurie H., Stryker, Sheldon. Theorizing the Relationship between Self-Esteem and Identity. In: Owens, T.J.Stryker, S., Goodman, N., editors. *Extending Self-Esteem Theory and Research: Sociological and Psychological Currents*. New York: Cambridge University Press; 2001. p. 29-55.
- Hadaway, C Kirk, Marler, Penny Long, Chaves, Mark. Overreporting Church Attendance in America: Evidence That Demands the Same Verdict. *American Sociological Review*. 1998; 63:122–30.
- Higgins, E Tory. Self-Discrepancy: A Theory Relating to Self and Affect. *Psychological Review*. 1987; 94:319–40. [PubMed: 3615707]
- Hogg, Michael A., Terry, Deborah J., White, Katherine M. A Tale of Two Theories: A Critical Comparison of Identity Theory with Social Identity Theory. *Social Psychology Quarterly*. 1995; 58:255–69.
- Holbrook, Allyson L., Krosnick, Jon A. Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique. *Public Opinion Quarterly*. 2010; 74:37–67.
- Jacobs, David R., Ainsworth, Barbara E., Hartman, Terry J., Leon, Arthur S. A Simultaneous Evaluation of 10 Commonly Used Physical Activity Questionnaires. *Medicine and Science in Sports and Exercise*. 1992; 25:81–91.
- James, William. *Principles of Psychology*. New York: Dover; [1890] 1950.
- Klesges, Robert C., Eck, Linda H., Mellon, Michael W., Fulliton, William, Somes, Grant W., Hanson, Cindy L. The Accuracy of Self-Reports of Physical Activity. *Medicine and Science in Sports and Exercise*. 1990; 22:690–97. [PubMed: 2233209]
- Kreuter, Frauke, Presser, Stanley, Tourangeau, Roger. Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly*. 2008; 72:847–65.
- Large, Michael D., Marcussen, Kristen. Extending Identity Theory to Predict Differential Forms and Degrees of Psychological Distress. *Social Psychology Quarterly*. 2000; 63:49–59.
- Larson, Reed, Csikszentmihalyi, Mihaly. The Experience Sampling Method. In: Reis, HT., editor. *New Directions for Methodology of Social and Behavioral Science*. Vol. 15. San Francisco: Jossey-Bass; 1983. p. 41-56.
- Leenders, Nicole YJM., Michael Sherman, W., Nagaraja, HN., Lawrence Kien, C. Evaluation of Methods to Access Physical Activity in Free-Living Conditions. *Medicine and Science in Sports and Exercise*. 2001; 33:1233–40. [PubMed: 11445774]
- Marcussen, Kristen, Large, Michael D. Identities, Self-Esteem, and Psychological Distress: An Application of Identity-Discrepancy Theory. *Sociological Perspectives*. 2006; 49:1–24.
- Marcussen, Kristen, Large, Michael D. Using Identity Discrepancy Theory to Predict Psychological Distress. In: Burke, PJ.Owens, TJ.Serpe, RT., Thoits, PA., editors. *Advances in Identity Theory and Research*. New York: Kluwer/Plenum; 2003. p. 151-66.
- Matthews, Charles E., Freedson, Patty S. Field Trial of a Three-Dimensional Activity Monitor: Comparison with Self-Report. *Medicine and Science in Sports and Exercise*. 1995; 27:1071–78. [PubMed: 7564975]
- McCall, George J., Simmons, JL. *Identities and Interactions: An Examination of Human Associations in Everyday Life*. New York: Free Press; 1978.
- Mead, George H. *Mind, Self and Society*. Chicago: University of Chicago Press; 1934.
- Niemi, Ilris. Systematic Error in Behavioral Measurement: Comparing Results from Interview and Time Budget Studies. *Social Indicators Research*. 1993; 30:229–44.
- Patterson, Patricia. Reliability, Validity, and Methodological Response to the Assessment of Physical Activity Via Self-Report. *Research Quarterly for Exercise and Sport*. 2000; 71:15–20. [PubMed: 25680008]
- Pew Research Center. From Telephone to the Web: The Challenge of Mode of Interview Effects in Public Opinion Polls. 2015a. Retrieved May 21, 2015 ([http://www.pewresearch.org/files/2015/05/2015-05-13\\_mode-study\\_REPORT.pdf](http://www.pewresearch.org/files/2015/05/2015-05-13_mode-study_REPORT.pdf))

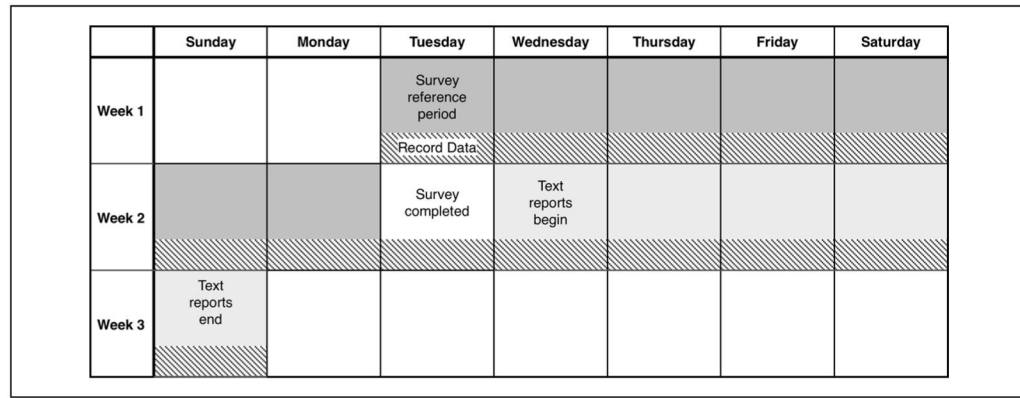
- Pew Research Center. The Smartphone Difference. 2015b. Retrieved May 20, 2015 (<http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015>)
- Presser, Stanley. Can Changes in Context Reduce Vote Overreporting in Surveys? *Public Opinion Quarterly*. 1990; 54:586–93.
- Presser, Stanley, Stinson, Linda. Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance. *American Sociological Review*. 1998; 63:137–45.
- Robinson, John P. The Validity and Reliability of Diaries Versus Alternative Time Use Measures. In: Juster, FT., Stafford, FP., editors. *Time, Goods, and Well-Being*. Ann Arbor, MI: Institute for Social Research; 1985. p. 33-62.
- Robinson, John P. The Time-Diary Method: Structure and Uses. In: Pentland, WE, Harvey, AS, Powell Lawton, M., McColl, MA., editors. *Time Use Research in the Social Sciences*. New York: Kluwer/Plenum; 1999. p. 47-90.
- Rosenberg, Morris. *Conceiving the Self*. New York: Basic; 1979.
- Rzewnicki, Randy, Auweele, Yves, Vanden, De Bourdeaudhuij, Ilse. Addressing Overreporting on the International Physical Activity Questionnaire (IPAQ) Telephone Survey with a Population Sample. *Public Health Nutrition*. 2003; 6:299–305. [PubMed: 12740079]
- Schuman, Howard. Artifacts Are in the Eye of the Beholder. *American Sociologist*. 1982; 17:21–28.
- Schwarz, Norbert. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*. Mahwah, NJ: Erlbaum; 1996.
- Schwarz, Norbert, Oyserman, Daphna, Peytcheva, Emilia. Cognition, Communication, and Culture: Implications for the Survey Response Process. In: Braun, M, Edwards, B, Harkness, J, Johnson, T, Lyberg, L, Mohler, P, Pennell, BE., Smith, TW., editors. *Survey Methods in Multinational, Multiregional, and Multicultural Context*. Hoboken, NJ: Wiley; 2010. p. 177-90.
- Serpe, Richard T. Stability and Change in the Self: A Structural Symbolic Interactionist Explanation. *Social Psychology Quarterly*. 1987; 50:44–55.
- Shephard RJ. Limits to the Measurement of Habitual Physical Activity by Questionnaires. *British Journal of Sports Medicine*. 2003; 37:197–206. [PubMed: 12782543]
- Stets, Jan E., Burke, Peter J. Identity Theory and Social Identity Theory. *Social Psychology Quarterly*. 2000; 63:224–237.
- Stets, Jan E., Burke, Peter J. New Directions in Identity Control Theory. *Advances in Group Processes*. 2005; 22:43–64.
- Stets, Jan E., Serpe, Richard T. Identity Theory. In: DeLamater, J., Ward, A., editors. *Handbook of Social Psychology*. 2. New York: Springer; 2013. p. 31-60.
- Stinson, Linda L. Measuring How People Spend Their Time: A Time Use Survey Design. *Monthly Labor Review*. 1999; 122:12–19.
- Stocké, Volker. Response Privacy and Elapsed Time Since Election Day as Determinants for Vote Overreporting. *International Journal of Public Opinion Research*. 2007; 19:237–46.
- Stryker, Sheldon. *Symbolic Interactionism: A Social Structural Version*. Caldwell, NJ: Blackburn; 1980.
- Stryker, Sheldon, Burke, Peter J. The Past, Present, and Future of an Identity Theory. *Social Psychology Quarterly*. 2000; 63:284–297.
- Stryker, Sheldon, Serpe, Richard T. Identity Salience and Psychological Centrality: Equivalent, Overlapping, or Complimentary Concepts? *Social Psychology Quarterly*. 1994; 57:16–35.
- Sudman, Seymour, Bradburn, Norman M., Schwarz, Norbert. *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass; 1996.
- Swann, William B. Self-Verification: Bringing Social Reality into Harmony with the Self. In: Suls, J., Greenwald, AG., editors. *Social Psychological Perspectives on the Self*. Vol. 3. Hillsdale, NJ: Erlbaum; 1983. p. 33-66.
- Swann, William B., Pelham, Brett W., Krull, Douglas S. Agreeable Fancy or Disagreeable Truth? Reconciling Self-Enhancement and Self-Verification. *Journal of Personality and Social Psychology*. 1989; 57:782–91. [PubMed: 2810025]
- Terry, Deborah J., Hogg, Michael A. Group Norms and the Attitude-Behavior Relationship: A Role for Group Identification. *Personality and Social Psychology Bulletin*. 1996; 22:776–93.

- Terry, Deborah J., Hogg, Michael A., White, Katherine M. The Theory of Planned Behavior: Self-Identity, Social Identity, and Group Norms. *British Journal of Social Psychology*. 1999; 38:225–44. [PubMed: 10520477]
- Tourangeau, Roger, Yan, Ting. Sensitive Questions in Surveys. *Psychological Bulletin*. 2007; 133:859–83. [PubMed: 17723033]
- Uskul, Ayse, Oyserman, Daphna, Schwarz, Norbert. Cultural Emphasis on Honor, Modesty, or Self-Enhancement: Implications for the Survey Response Process. In: Braun, M.Edwards, B.Harkness, J.Johnson, T.Lyberg, L.Mohler, P.Pennell, BE., Smith, TW., editors. *Survey Methods in Multinational, Multiregional, and Multicultural Context*. Hoboken, NJ: Wiley; 2010. p. 191-202.
- Wellek, Stefan. *Testing Statistical Hypotheses of Equivalence and Noninferiority*. 2. Boca Raton, FL: Taylor and Francis; 2010.
- White, Katherine M., Terry, Deborah J., Hogg, Michael A. Safer Sex Behavior: The Role of Attitudes, Norms, and Control Factors. *Journal of Applied Social Psychology*. 1994; 24:2164–92.
- Zuzanek, Jiri, Smale, Bryan JA. Life-Cycle and Across-the-Week Allocation of Time to Daily Activities. In: Pentland, WE.Harvey, AS.Powell Lawton, M., McColl, MA., editors. *Time Use Research in the Social Sciences*. New York: Kluwer/Plenum; 1999. p. 127-54.

## Biographies

**Philip S. Brenner** is an assistant professor in the Department of Sociology and senior research fellow in the Center for Survey Research at the University of Massachusetts Boston. His current research investigates identity processes in survey data collection. His work has recently appeared in *Social Psychology Quarterly*, *Sociological Methods & Research*, *Social Forces*, and *Public Opinion Quarterly*.

**John DeLamater** is a Conway-Bascom Professor of Sociology at the University of Wisconsin, Madison. His research brings a broad, biopsychosocial perspective to the study of sexuality through the life course. His research has appeared in the *Journal of Marriage and Family*, *Journal of Sex Research*, *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, and *Social Indicators Research*. He has published papers for more than 40 years on the quality of self-reports in social psychological research.



**Figure 1.**  
Example of Study Sequence by Reference Periods

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

		<b>S = 0</b> Self-reported nonexercisers	<b>S &gt; 0</b> Self-reported exercisers
<b>C = 0</b> Nonexercisers		S = C Admitted nonexerciser	S > C Nonexercising overreporter
<b>C &gt; 0</b> Exercisers		S > C Underreporter	S > C Exercising overreporter
			S = C Validated exerciser

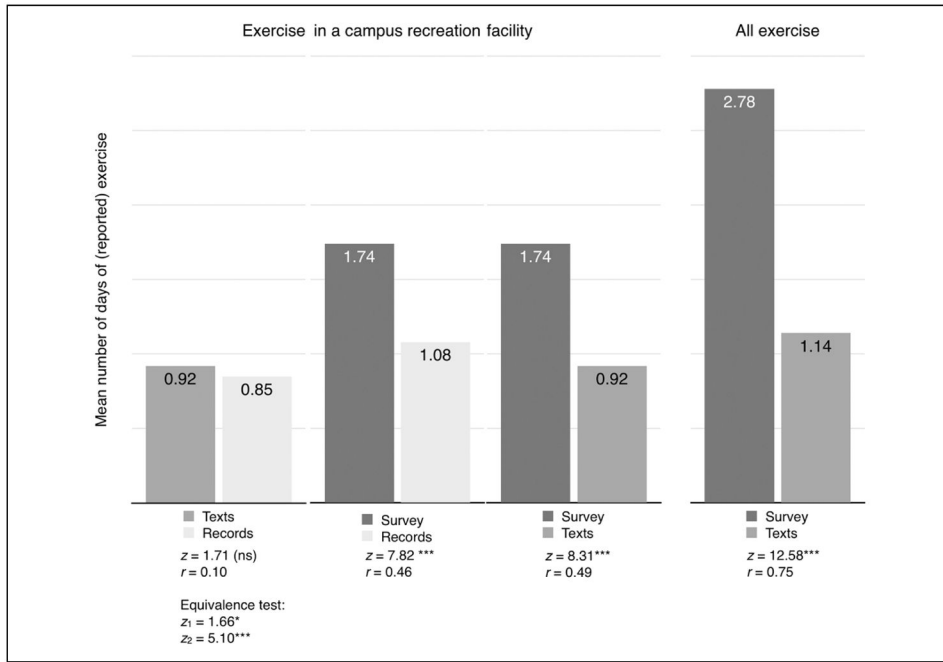
**Figure 2.** Potential Outcomes from Comparison between Survey Response (S) and Criterion Measures (C)

Author Manuscript

Author Manuscript

Author Manuscript

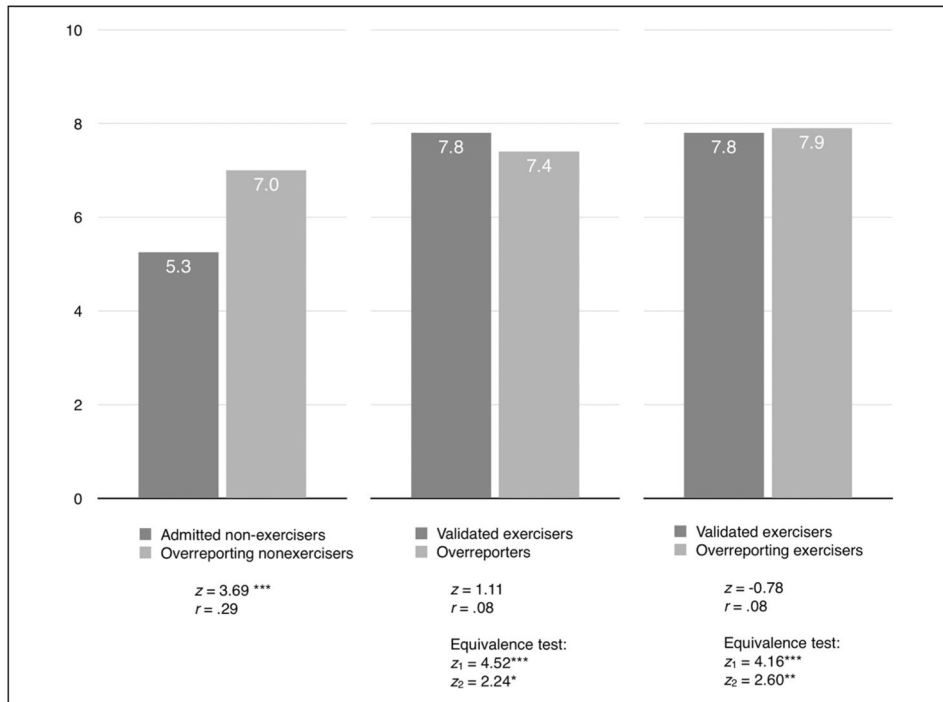
Author Manuscript



**Figure 3.** Comparisons of Mean Number of Days of Exercise and Reported Exercise by Data Collection Methods and Location of Exercise

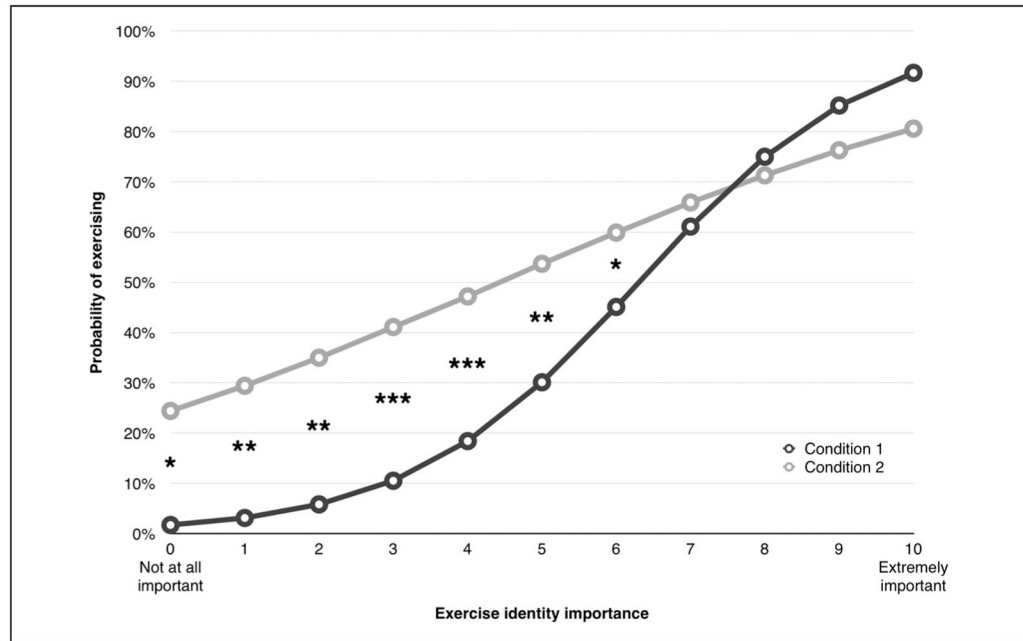
*Note:* Tests (Wilcoxon signed-rank) use paired data; equivalence tests use TOST signed-rank. The survey and text measures are from sequential reference periods, not the identical reference period. All other comparisons are from matching reference periods.

\* $p < .05$ . \*\*\* $p < .001$ .



**Figure 4.** Comparison of Exercise Identity Prominence Ratings by Respondent Type  
*Note:* Tests use Mann-Whitney-Wilcoxon rank-sum. Equivalence tests use the TOST rank-sum. The survey and text measures are from sequential reference periods, not the identical reference period. All other comparisons are from matching reference periods.  
 \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .





**Figure 5.**  
 Predicted Probability of Exercising by Condition and Exercise Identity Importance  
 \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

**Table 1**

Predicting Exercise with Condition, Identity Prominence, and the Interaction of Condition and Identity Prominence

	Texted reports of exercise								
	Condition only			Main effects only			Full model		
	Coefficient	SE	<i>p</i>	Coefficient	SE	<i>p</i>	Coefficient	SE	<i>p</i>
<b>Condition (directive = 1)</b>	.222	.242		.346	.267		2.957	1.016	**
<b>Identity prominence</b>				.397	.064	***	.648	.125	***
<b>Interaction</b>							-.393	.146	**

\*\*  
*p* < .01.

\*\*\*  
*p* < .001.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript