

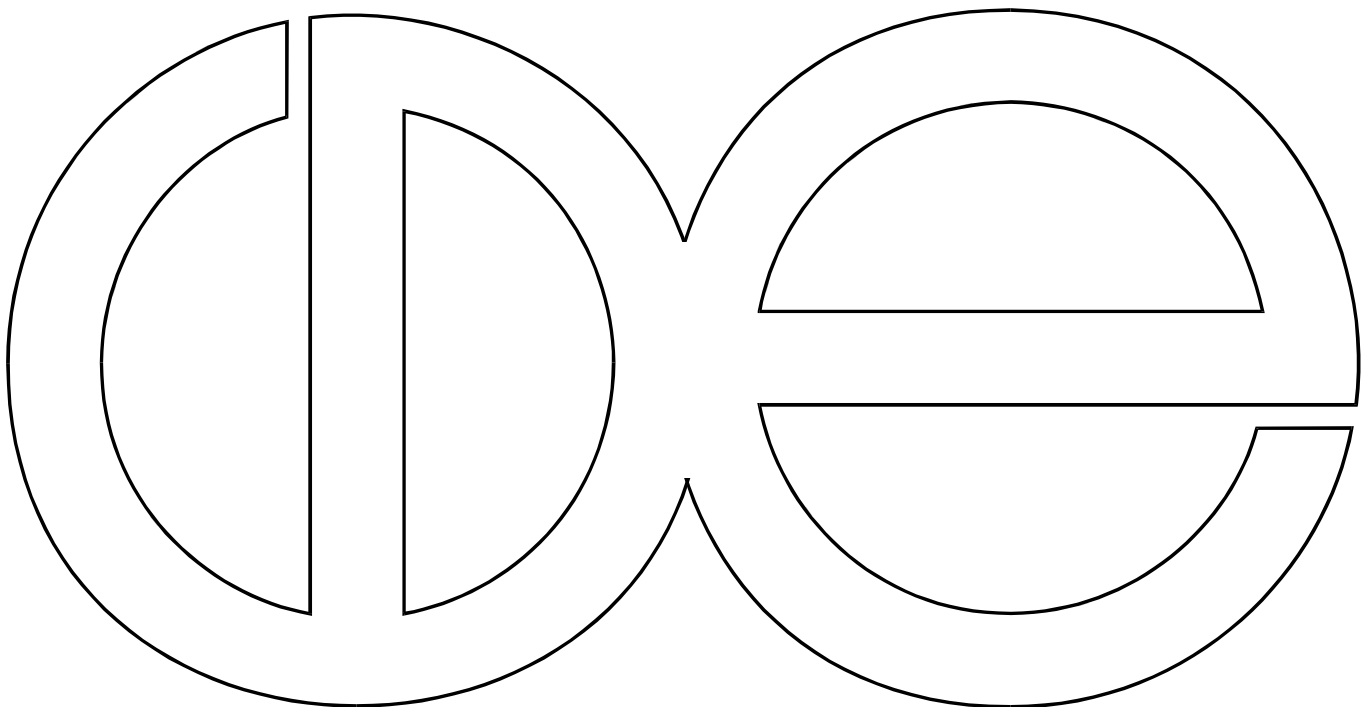
Center for Demography and Ecology
University of Wisconsin-Madison

**Overreporting of Exercise in a Self-administered Mode:
The Biasing Effect of Identity on Survey Responses**

Philip S. Brenner

John D. DeLamater

CDE Working Paper No. 2013-02



Overreporting of Exercise in a Self-administered Mode: The Biasing Effect of Identity on Survey Responses

Philip S. Brenner
University of Massachusetts, Boston

John D. DeLamater
University of Wisconsin, Madison

Abstract

Like that of other socially desirable behaviors, much of the research on physical exercise is based on self-reports which are prone to overreporting. While research has focused on identifying the presence and degree of overreporting, this paper fills an important gap by investigating its causes. The explanation based in impression management will be challenged, using an explanation based in identity theory as an arguably better fitting alternative. Respondents were randomly assigned to one of two conditions: (1) a web instrument using direct survey questions, or (2) a chronological reporting procedure using text messaging. Comparisons to validation data from a reverse record check indicate significantly greater rates of overreporting in the web condition than in the text condition. Results suggest that measurement bias is associated with the importance of the respondents' exercise identity, prompted by the directness of the conventional survey question. Findings call into question the benefit of self-administration for bias reduction in measurement of normative behaviors.

The relationship between physical activity and health is positive and strong. Therefore, policies that advocate health-promoting behaviors, like exercise, are seen as very important for encouraging the physical health of individuals in our society (US Department of Health and Human Services 2008). However, good health policy decisions depend on accurate information, and accurate information depends on good measures. According to Washburn et al., “the task of determining effective interventions for promoting adoption and maintenance of healthful physical activity behaviors will be impossible without reliable and valid measurements of the dependent variable, physical activity” (2000, p. 105). Accordingly, the accurate measurement of these activities has become essential for effective policy creation and implementation.

Much of the extant research on the relationship between physical activity and health is based on self-reports, typically gathered through survey research. However, similar to other normative behaviors like church attendance (see Chaves and Stephens 2003 for a review) and voting (Bernstein, Chadha, and Montjoy 2001), physical exercise is often overreported (Sallis and Saelens 2000). Respondents report higher rates or more frequent activity on surveys than actual behavior warrants, leading questionnaire measures of physical activity to suffer from low validity (Durante and Ainsworth 1996; Shephard 2003).

Epidemiologists have tested the validity of questionnaire measures of physical activity using objective physical measures as a comparator. These objective physical comparators have included percentage of body fat (Ainsworth, Jacobs, and Leon 1992), exercise performance data recorded by equipment like treadmills, accelerometers and actigraphs (Adams et al. 2005; Jacobs et al. 1992; Leenders et al. 2001; Matthews et al. 2000; Matthews and Freedson 1995), an estimation of metabolic rate using doubly labeled water (Adams et al. 2005), and an estimation

of energy expenditure from the caloric intake of a controlled feeding program (Albanes et al. 1990). These comparators, each with its own validity problems (see Bassett 2000; Williams et al. 1989) may be better described as evidence of convergent rather than criterion validity (Patterson 2000; see also Washburn, Heath, and Jackson 2000) as none compare a self-report of physical exercise with validation data for the same activity. Rather, each compares a survey self-report with potential evidence, of varying quality, of that activity.

Other work has utilized methods and data more common to the social sciences, like flexible interviewing techniques, direct observation, and reverse record checks to estimate the validity of self-reports. Suspecting a high amount of overreporting, Rzewnicki et al. (2002) reinterviewed a random subsample of survey respondents using a more detailed set of probes. The reinterview uncovered high levels of overreporting that were attributed to problems in interviewing and question comprehension in the initial interview (see also Baranowski 1988; Durante and Ainsworth 1996). Using direct but unobtrusive observation, Klesges et al. (1990) found that respondents who reported high levels of habitual physical activity showed the greatest level of divergence between their self-reported and actual levels of activity during the reference period. Chase and Godbey (1983) asked members of a tennis club and a swimming club to estimate the frequency of their past use of the club's facilities and compared these self-reports to sign-in sheets. In both clubs, they found that over 75% of respondents overestimated their participation in club activities.

While these projects have been very informative in their primary focus on identifying the presence and degree of overreporting, very few studies have attempted to explain its causes. Comparing objective physical measures (e.g., accelerometer data, estimates of energy

expenditure) to a series of physical activity recalls, Adams et al. (2005) tested the Martin-Larson Approval Motivation Scale and the Crowne-Marlow Social Desirability Scale as predictors of overreporting. They found that social approval was not strongly associated with overreporting and that the relationship between social desirability and overreporting was inconsistent. Motl, McAuley, and DiStefano (2005) too found only a minimal influence of social desirability on self-reported physical activity. However, their lack of a criterion measure limited analysis to self-reports rather than to estimates of overreporting, adding a major caveat to their inferences.

In summary, while the existence of the phenomenon is well established, very little research has explored the causes of overreporting of physical exercise on surveys, and the literature that does exist fails to consistently and persuasively explain why overreporting occurs. Moreover, previous research has found that the hypothesized positive relationship between overreporting and social desirability is either weak and inconsistent, or nonexistent. This project aims to fill in this gap in the extant literature by investigating the causes of overreporting of physical activity. In so doing, the currently accepted explanation of social desirability effects for the overreporting of normative behaviors based in impression management theory will be challenged, using an explanation based in identity theory (Stryker 1980/2003) as an arguably better fitting alternative.

Social desirability effect: impression management or an identity process?

The cognitive model of survey response (Cannell, Marquis, and Laurent 1977; Cannell, Miller, and Oksenberg 1981; Tourangeau 1984; Tourangeau, Rips and Rasinski 2000) established a set of steps for the response process, allowing survey methodologists to further pursue the “total survey error” approach (Groves 1989/2004; Weisberg 2005) and identify the sources of

error in the survey process. The final two stages of the process, judgment and response, are generally viewed as those in which social desirability effects emerge. That is, social desirability considerations result in the intentional misreporting of sensitive behaviors. For example, respondents tend to underreport contranormative, sensitive, or embarrassing behaviors (e.g., illicit drug use and drunk driving arrests) and overreport normative or positively viewed behaviors (e.g., voting, church attendance, and physical exercise). In its common formulation, the respondent misreports behavior in order to manage the impression s/he is presenting to the interviewer. Therefore, removing the interviewer from the measurement process through the use of self-administered questionnaires is generally accepted as a means to achieve more valid reports of sensitive behaviors (Aquilino and LoSciuto 1990).

While evidence in the literature is strongly supportive of the impression management paradigm for the underreporting of contranormative behavior (i.e., self-administered modes result in higher, presumably valid, estimates) comparatively little research has tested its application to the overreporting of normative behaviors. If impression management is driving overreporting of normative behavior in interviewer-administered modes, we should see a reduction in overreporting when comparing interviewer- and self-administered modes from comparable samples (e.g., random subsamples) of respondents. However, the few existing tests of this assumption either fail to support this approach or are inconclusive (Kreuter, Presser, and Tourangeau 2008; see Stocké 2007 for an exception.)

An alternative explanation focuses on the respondent's identity. Identity theory (Stryker 1980) contributes two concepts that are useful for understanding the difference between these two modes of survey measurement. Importance is the "personal value individuals place on an

identity (that) taps into subjective feelings of what is central to individuals' conceptions of themselves" (Ervin and Stryker 2001:34-5). While a respondent is aware of the importance s/he places on an identity, s/he does not necessarily know how likely s/he is to perform that identity. This second concept of identity theory, salience, is the probability of an identity being invoked within a situation (Stryker 1980) or the propensity to define a situation in a way as to provide an opportunity to perform that identity (Stryker and Serpe 1982). The salience of an identity indicates its relative position in a hierarchy of identities in which each identity is ranked according to its propensity of being situationally "called up."

While self-reported levels of importance and salience of an identity are often consistent, they need not be. Congruence may be an artifact of measurement (Stryker and Serpe 1994). The process of measuring the salience of an important identity may inflate the self-report of salience to match the importance of the identity. Moreover, the directive nature of survey measurement makes it difficult to conceive of a measure of identity salience that does not prime the identity, encouraging the alignment of importance and salience, and affecting the process of measurement (Ervin and Stryker 2001; Stryker and Burke 2000). Such a situation is especially problematic for the measurement of an important, but rarely enacted, identity (Burke 1980).

However, unlike conventional survey questions, chronologically based data collection procedures, like time diaries, avoid much of the biasing effect of identity importance (Bolger, Davis, and Rafaeli 2003). By avoiding direct questions about specific behaviors of interest (Robinson 1985, 1999; Stinson 1999), chronological data collection procedures avoid prompting self-reflection on the part of the respondent, arguably yielding less biased and higher quality data on many normative behaviors (Bolger et al. 2003; Niemi 1993; Zuzanek and Smale 1999).

The following analyses will predict overreporting of physical exercise, computed as the difference between self-reported and “gold standard” measures of behavior, using a measure of identity importance and measurement condition (i.e., the measurement of salience using either a direct survey question or a nondirective chronological reporting procedure) as predictors. The gold standard validation measure — a reverse record check of admission records to recreation facilities¹ — will be treated here as the true value of salience; the actual propensity of identity enactment during the reference period. The two self-reported measures (which will be described shortly) are observations of this true value of varying quality. A conventional survey self-report is hypothesized to be a biased measure of the true value with the bias arguably caused by identity importance. However, the experimental chronological reporting procedure is expected to produce a relatively unbiased measure of the true value. Including both identity and method of measurement provides a test of a main assumption of this application of identity theory to survey responding on normative behavior.

Data and methods

A sample of undergraduates from a large, Midwestern university were invited to participate in a survey of daily life conducted in March and April 2011. While certainly a convenient population, more than their accessibility made students an ideal sampling frame. Similar to other American universities, students at the selected university automatically have access to the university’s recreation facilities. Sampling from the list of enrolled students from

¹ Direct but unobtrusive observation was carried out at recreation facilities to ensure the quality of these record data. During the periods of observation, all of the individuals who entered the recreation facilities had their cards scanned, and all then proceeded to engage in exercise activities or sports, although some first stopped briefly to change into workout clothes in a locker room. In short, these observations reassured the researchers that the record data are a valid measure of exercise and sport activities.

the bursar allowed the researchers to sample gym members without the selection biases that one would encounter if a sampling frame of the membership rolls of for-profit gym, or non-profit or community fitness center (e.g., YMCA, city or county owned recreation center) was utilized.

A random sample of 600 undergraduates, stratified by gender and year in school, was divided into three subsamples: 250 were invited to take a web survey, 250 were invited to take a similar web survey followed up by five days of data collection via SMS/text messaging; 100 were reserved as a replicate, later employed with 25 in the web-only condition and 75 in the text condition based on the achieved number of responses at the time. The response rate varied between conditions. 124 respondents completed the web questionnaire in each condition, yielding an overall response rate² of 41.3 percent; 45.1 percent for the web-only condition, and 38.2 percent for the text condition.

An invitation to participate in the study was sent to the student's university email address, including a link to the web survey. An email reminder to complete the survey was sent three days after the initial invitation, and a final reminder was sent five days after the first reminder email. Participants in both conditions were offered a ten dollar incentive following completion of the web survey. Participants in the text condition were offered an additional 30 dollars following completion of the text component.

The web surveys in both conditions were comprised of approximately twenty questions about usage of university facilities. While the true purpose of the study was to measure use of university recreation facilities, questions about the use of campus libraries, the student union, and other facilities were also asked in order to mask the focus of the study. Respondents in both

² All response rates are computed as AAPOR RR 5; there were no ineligible cases nor any cases of unknown eligibility.

conditions were asked about their “typical” use of recreational facilities on campus in terms of their “usual” activities at these facilities (e.g., weightlifting, swimming, aerobics, and cross-training) and the time they spend on these activities. These typical and usual exercise questions were asked to allow respondents to claim an exercise identity to reduce the normative bias on the actual question of interest. Respondents were then asked to report actual behavior for a short, well-defined reference period; the method of self-report of actual behavior differed between conditions/survey modes. Respondents in the web-only condition were asked to report the number of days in the past seven in which they exercised and on how many of these days, if any, their exercise activity occurred at a campus recreational facility.

Respondents in the text condition, in lieu of being asked the direct survey question, were asked to send SMS/text messages to the research team reporting all changes in their major activities for a period of five days. Respondents received training documents detailing how and what to report and were reminded multiple times each day to send messages updating their activities. These reminders were more frequent on the first day of their participation (four reminders, at 10:00 am, 1:00 pm, 5:00 pm, and 8:00 pm) and less frequent on the final days of participation (two reminders, at 10:00 am and 8:00 pm). The frequency and other characteristics of the text message activity updates suggest that these data are of high quality (Brenner and DeLamater 2012). Respondents in the text condition were asked to report all changes in the major daily activities and where they were taking place without reference to physical exercise or sports.

After all other participation was completed, respondents in both conditions were asked for their student identification numbers so that study staff could request records on their use of

campus recreation facilities. These records are the product of the swiping of students' identification cards upon admission to the facilities. This process records the student's identification number and the time and day of admittance to the facility. Of the 124 web-only respondents, 89 granted access to their record data, yielding a compliance rate of 71.8 percent and a final response rate in this condition of 32.4 percent. Of the 87 respondents in the text mode who completed the texting component, 67 permitted access to their record data, yielding a compliance rate of 77.0 percent and final response rate of 22.8 percent in this condition.

Measures.

Three key variables are used in the following analyses. The first is a measure of the validity of the self-report, computed as the difference between the reverse record check and the respondent's self-report. Each day with a record of admittance during the reference period was coded as 1, 0 otherwise. This procedure yielded a series of variables, one for each day, each coded for the presence or absence of an admittance. These were summed to reflect the number of days during the reference period that the respondent used campus recreation sports facilities.

This measure differs between conditions. For the web-only condition, the record check variable was compared to the direct survey question asking about use of campus recreational sports facilities during the reference period. This question was preceded with the more general question, "In the past seven days, how many days have you worked out or exercised?" The intent of this more general question (as well as the "typical" use questions previously discussed) was to achieve less biased measurement of campus recreational sports facility use by allowing the respondent to claim exercise behavior even if it did not occur at campus recreational sports

facilities before asking the more specific question of interest: “Of the days you worked out in the past week, how many did you use [University Name] Recreational Sports facilities, like [names of facilities]?” The result of the reverse record check was subtracted from the self-report to create the measure of the validity of the self-report.

In the text condition, respondents reported changes in their major activities during the reference period using SMS/text messages. These messages were coded for exercise activities and, more specifically, for those that occurred at campus recreational sports facilities. Each day with a report of exercise at a campus recreation sports facility was coded as 1, 0 otherwise. This variable was then summed over the days of the reference period. As in the web condition, the difference between this self-report and the record variable provided an estimate of overreporting. This procedure resulted in a nominal variable: (1) validated exercisers, (2) admitted non-exercisers, and (3) overreporters.³ This variable is used as a dependent variable in the first analysis, and a grouping variable in the second analysis.

The second variable is a measure of the importance of a physical exercise identity. In the initial web survey, all respondents were asked a battery of eight questions about the importance of a set of identities common amongst college students. “Each of us is involved in various activities. How important to you is: playing sports, exercising, or working out?” Identity importance was measured using a five-point scale from “not at all important” to “extremely important.” Nearly two-thirds of the respondents reported that an exercise identity was very or extremely important. This variable is a key variable the second analysis.

³ Some amount of underreporting is inevitable as respondents forget to report activities. The rate of underreporting does not differ between conditions (10 and 4 percent in the text and web modes, respectively; $X^2(1) = 2.06$; $p=0.15$), suggesting that underreporting is not attributable to the characteristics of the mode. As is not a focus of this study, it will not be discussed further here.

The final variable is a dichotomous indicator of the treatment (directness of the measure) condition: the conventional, direct survey question in the web-only condition, coded 1, and the indirect measurement in the text condition, coded 0.

Analysis plan

Two hypotheses will be tested. First, the difference in the rate of overreporting between conditions will be examined. Overreporting is hypothesized to be less likely to occur in the text condition compared to the web condition, as the text method lacks a direct survey question. Differences in the rates of overreporting between the two conditions will be tested for significance using a chi-square test.

Second, the effect of importance will be examined between subgroups of respondents. The first comparison focuses on respondents who report having exercised on the web-survey. If identity importance prompts an affirmative self-report, regardless of actual behavior, no difference in importance ratings should emerge between the two groups of self-reported exercisers: validated exercisers and overreporters. Alternatively, if the misreport is the result of a stochastic process, a difference between these groups should emerge. That is, if the misreport is caused by a random process orthogonal to identity (e.g., random misremembering), a distinct difference should emerge between overreporters and actual exercisers.

The second comparison focuses on respondents whose records indicate that they did not visit campus recreation facilities during the reference period. If identity importance prompts the self-report of exercise, a difference in the reported importance of an exercise identity should emerge when the two groups of non-exercisers are compared: overreporters should report a

higher level of identity importance compared to admitted non-exercisers. Alternatively, if the misreport is the result of a stochastic process (e.g., random forgetting), no difference between these two groups should emerge. The Mann-Whitney-Wilcoxon (MWW) test will be used to compare identity importance ratings across subgroups.

Results

As hypothesized, overreporting emerged in the web condition and at a much higher rate than in the text condition (see Table 1). Almost half (43/89) of the respondents in the web-only condition overreported their frequency of exercise. Moreover, of the 63 web respondents who claimed exercise in the past week, over two-thirds overreported. Overreporting in the text condition, however, was negligible. Only five text respondents (of 67) overreported their exercise. Whether the rate of overreporting is compared for all respondents, or only for self-reported exercisers, the difference between conditions is highly statistically significant ($p \leq 0.001$; $X^2 = 28.1$ and 16.8 respectively, with 1 d.f.).

[Table 1 about here]

Similar analyses compared validated exercisers and admitted non-exercisers between conditions. Respondents did not differ in their likelihood to accurately claim having exercised between conditions ($X^2 = 2.07$; n.s.). However, respondents in the text condition were more likely than those in the web condition to accurately report not exercising ($X^2 = 10.1$; $p \leq 0.001$).

Given the lack of substantive and significant overreporting in the text condition, the remaining comparisons will focus only on respondents in the web condition. The first of these compares the identity importance ratings of the two groups of self-reported exercisers:

overreporters and validated exercisers. Both of these groups of respondents rate the importance of their exercise identity highly ($\bar{Y}_{ve} = 4.4$, $\bar{Y}_o = 4.2$). Overreporters and validated exercisers rate the importance of their exercise identity similarly, from the highest (40 and 42 percent, respectively) to lowest (2 and 0 percent) levels (see Figure 1). Using an MWW test, the null hypothesis of equality of the distributions of identity ratings between validated exercisers and overreporters cannot be rejected ($z = 1.04$; n.s.; see Table 2).⁴

[Figure 1 about here]

[Table 2 about here]

The second comparison looks for a difference in the identity importance rating between overreporters and admitted non-exercisers (see Table 2). Admitted non-exercisers rate the importance of their exercise identity much lower ($\bar{Y}_{ane} = 3.7$) than do overreporters ($\bar{Y}_o = 4.2$). More overreporters rate the importance of their exercise identity highly than do admitted non-exercisers (40 and 26 percent, respectively) and fewer rate it less important (2 and 11 percent, respectively). Using MWW, the null hypothesis of the equality of these distributions is rejected ($z = -2.07$; $p \leq 0.05$). While this finding supports the hypotheses based in identity theory, some further analysis may be enlightening.

[Figure 2 about here]

Overreporters, as currently defined, are not necessarily non-exercisers. Rather, a number of respondents in this group (19 of 43) actually engaged in exercise but overreported its frequency. A more apt comparison would be one juxtaposing the two groups of non-exercisers: non-exercising overreporters and admitted non-exercisers. Non-exercising overreporters ($\bar{Y}_{neo} =$

⁴ A t-test comparing these means comes to the same conclusion. Comparisons using t-tests match the results of all subsequent Mann-Whitney-Wilcoxon tests.

4.4) rate the importance of their exercise identity even more highly than the general group of overreporters. Non-exercising overreporters are about twice as likely to rate their exercise identity as extremely importance compared to admitted non-exercisers (50 and 26 percent, respectively) and less likely to rate as unimportant (0 and 11 percent, respectively). Again using MWW, the null hypothesis of equality of these distributions is rejected ($z = -2.66; p \leq 0.01$).

[Figure 3 about here]

Discussion

The first hypothesis, that the direct measure of exercise would promote overreporting, was strongly supported. Respondents in the web condition with the direct survey question were much more likely to overreport the number of days they exercised than were respondents in the nondirective text condition. Conversely, chronological reporting procedures, like time diaries, allow measurement of some normative behaviors without the bias inherent in direct survey question because they do not engage identities to the same extent. The SMS/text procedure used here, like the other chronological reporting modes, also appears to produce more valid, less biased, data on normative behaviors.

Moreover, this finding casts doubt on the assertion that, in general, self-administered modes lead to valid reports of normative behaviors. These findings suggest that a direct survey question, even in a self-administered mode, can lead to dramatic levels of overreporting. Nearly half of the respondents in the web mode overreported, compared to a negligible rate of overreporting in the text mode. While the current project only compares two self-administered modes and does not permit comparison to interviewer-administered data, it is difficult to imagine that the presence of an interviewer could produce substantially more biased responses than those

found in the web mode. Arguably, it is the direct survey question that, regardless of the mode in which it is administered, causes much of the bias in the measurement of normative behaviors. Clearly, further testing of this is an excellent opportunity for future research.

The second hypothesis, that identity importance is associated with overreporting, was also supported. According to identity theory, the importance of an identity can bias the measurement of its salience. The salience of an identity of high importance may be overreported if these two concepts become conflated during the measurement process (Stryker and Serpe 1994). It has been argued that direct survey measurement promotes this conflation of importance and salience (Brenner 2011; Burke 1980) and it appears that just such a situation may have occurred here. Ratings of exercise identity importance were higher for overreporters compared to admitted non-exercisers but did not differ from those of verified exercisers.

Survey questions about behavior typically call for the respondent to recall relevant information and then either enumerate (i.e., count instances of the behavior during a reference period) or estimate (i.e., apply a rate based on recent behavior recalled from memory, a notion of one's "typical" behavior, or using other easily accessible information about one's behavioral patterns) to arrive at an answer. The direct survey question in the web condition calls for enumeration using the past week as the reference period. The respondent is asked to search his or her memory for the particular activity of interest over the reference period, sum up the occurrences, and report an answer. However, multiple opportunities for error arise in this procedure. First, the respondent may honestly misremember which days s/he went to the campus recreation facilities or inaccurately sum these occurrences. These unintentional errors should be stochastic, with a mean of zero, and just as likely to yield an underreport as an overreport. Given

the results of this study, random response error is not likely to be the primary contribution to the total survey error.

Second, the respondent may correctly remember and sum these occurrences, but may choose to edit this actual rate of behavior. This type of motivated misreport would likely be caused by the more conventional notion of social desirability (e.g., “I want the interviewer to think that I’m a health-conscious person”). Alternatively, the respondent may either be unable to recall his/her actual behavior over the reference period, or unwilling to put forward the cognitive resources to generate an accurate count (i.e., cognitive mising), either situation yielding estimation rather than enumeration. In this third case, identity is a prime suspect as a cause for the misestimation.

While these findings are not definitive in teasing apart these two different, yet related, explanations, they do suggest that the third possibility is likely and that identity has a primary role to play in promoting an overreport. Statistical tests demonstrate the strong association between identity importance and overreporting, distinguishing overreporters from admitted non-exercisers but not from verified exercisers. Moreover, that overreporting emerges in a self-administered form without the presence of an interviewer suggests that there is something other than social desirability bias, as it is conventionally understood, occurring. In the measurement of normative behaviors, social desirability bias is arguably not based primarily in the impression management activities of the respondent vis-à-vis an interviewer, but rather in the management of the respondent’s own impression of him- or herself. When asked a direct question about the salience of an identity during a brief, well defined reference period, the respondent with high identity importance may overreport the salience of the identity, not in order to impress an

interviewer, but rather as a verbal performance consistent with the identity s/he values strongly. That this process occurs in a self-administered mode with a direct question but not in a self-administered mode using nondirective chronological measurement is strongly supportive of this conclusion.

The two measures of exercise activity also differ on the role of memory in reporting. The direct survey question in the web survey relies on retrospective reporting whereas the text condition uses *in situ* responding. Research by Belli et al. (1999) has suggested that some amount of overreporting of normative behaviors may be related to the effect of memory in the retrospective report. When asked about behavior that cannot be easily recalled and counted, or that which was never encoded into memory, the importance of the identity may influence the estimation procedure. The recall or estimation procedure that occurs during the process of answering a direct question enables the biasing effect of identity. In short, identity importance helps the respondent fill in the gaps in his or her memory with what the s/he “usually” or “typically” does, or perhaps more to the point, what identities the respondent values in him- or herself. According to Hadaway, et al., “overreporting may be generated by the respondent’s desire to report truthfully his or her identity ... and the perception that the (question about identity-consistent behavior) is really about this identity rather than actual (behavior).” (adapted from Hadaway et al., 1998:127). However, the chronological reporting procedure used in the text condition avoids this process. Future work should strive to identify ways to reduce the possibly biasing effects of memory by limiting the reference period and the focal behavior upon which the respondent is reporting.

A possible avenue for future research may be an examination of these two possible causes of overreporting (social desirability and identity) and their correlation with the two question answering processes discussed here (estimation and enumeration). It may be that identity is the primary cause of overreporting when respondents estimate the frequency of a normative behavior and that more conventional notions of social desirability are the primary cause of overreporting when respondents enumerate occurrences of a normative behavior. Unfortunately, these data cannot answer this question as there is no way to know for certain which process respondents used to generate an answer to the past week exercise question.

Limitations and future directions

The differing reference periods between the two conditions was an unavoidable limitation of the study's design. Reporting via text message for seven days would have been a much greater burden for the respondents. Conversely, a five day reference period for the stylized question in the web condition would have been awkward, as a week is a more "natural"⁵ time frame for respondents. The main problem with this difference is that it allows respondents in the web condition more opportunity to overreport than those in the text condition, thereby increasing the bias in web condition compared to the text condition simply due to the study design. However, the difference in overreporting between the two conditions was so extreme that the difference in reference period is unlikely to be a major cause. Nevertheless, future work should attempt to overcome this flaw in the study design by using identical reference periods.

The characteristics of the sample may also limit the findings of the study. While the sample size is relatively small and the sample composition (undergraduates at an elite public

⁵ There is, of course, nothing natural about a week. Unlike the day, month, and year, the week is perhaps the most socially-constructed unit of time. (Zerbavel 1989).

university) is likely quite different from the general population, both of these characteristics provide the study certain strengths and open the door to future work. The size of the sample allowed intensive data collection that would not have been possible with a large sample at current funding levels. However, given adequate funding, future work could take advantage of computing resources (e.g., an automated SMS-to-HTTP service) unavailable to the researchers on this project and would allow a much larger sample at a similar level of intensity of data collection. University undergraduates provided both a convenient sample as well as one with strong physical exercise/sport identities, providing a great resource for hypothesis testing using an experimental design. Students in highly selective universities may, for example, place a great deal more importance on their exercise identity, leading them to both exercise and overreport at higher rates than the general population. While this is a strength in the current work, allowing a strong test of the hypotheses, extending this framework to a general, non-student population would further test this model. Moreover, a sample from the general population may yield a more balanced distribution of exercise identity importance (including more respondents with low ratings of exercise importance), allowing a more complete test of these hypotheses.

Nonresponse may also present a caveat to inferences made from this study. Respondents were presented with multiple opportunities to decide on their (continued) participation: when initially recruited to participate, when asked for a cellular phone number, when asked to begin sending text updates on their activities (for text respondents), and when asked for their student identification number and permission to access their recreation facilities records. The students who chose to participate (repeatedly, given the multiple opportunities to opt out) may be different from those who chose not to participate at all or who opted out at various stages during the data

collection process. While it is likely that each additional step of informed consent could lead to pools of respondents who differ from nonrespondents and from each other, no differences emerge on information variables on the sampling frame (i.e., gender, year in school) or collected during at various stages (i.e., agreeing to participate in the text condition; granting permission to access recreation facilities records). These supplementary analyses (not shown) suggest that nonresponse may not be causing differential response patterns. Moreover, findings from investigations of overreporting of other normative behaviors suggest that those sample members who engage in normative behaviors, like attending religious services, are more likely to agree to a request for survey participation (see Woodberry 1998). If respondents differ from nonrespondents, they should more likely be actual exercisers rather than overreporters. In short, any nonresponse bias is likely to make these estimates of overreporting more conservative.

Conclusion

The findings of this study add credence to an identity approach to the study of overreporting of normative behavior. This approach moves the locus of the misreport from the interaction between the interviewer and the respondent, an understanding based in impression management theory, to the internal conversation the respondent has with him- or herself. The motivation to present oneself in a manner consistent with important identities can lead to overreporting on direct survey questions, even in self-administered modes. In essence, the survey question becomes an opportunity to verbally perform a highly important identity even if it is rarely performed in other situations (Burke 1980). However, the lack of an opportunity to perform an important identity in chronological reporting procedures, like time diaries and the

SMS/text procedure introduced here, yields more accurate reports by preventing the priming of an identity and avoiding the bias it elicits.

This paper extends current work investigating the quality of measurement of normative behaviors, using physical exercise as a case for study. Focusing on the measurement of health behaviors is ideal not only because it provides a set of potentially verifiable normative behaviors that can be compared to self-reports but also because of its importance to policymakers, physicians, and other stakeholders. By comparing respondents who accurately report their behavior with those who overreport, and theorizing about important contributing factors, we can gain insight into (1) what encourages positive health behaviors like exercising or getting regular medical checkups, (2) individuals' assessments of their own health and health behaviors, as well as (3) the causes of survey misreporting more generally.

This approach implies that measurement error is not just an annoyance to be avoided. Rather, as Howard Schuman (1982) suggested, these errors provide us an excellent opportunity to learn about ourselves. Studying misreporting can have real-world implications, especially in the realm of health policy. When survey respondents misreport their health behaviors, like physical exercise, they are tapping into a part of themselves they see as important. Going beyond the "lie" and gaining a better understanding of the reasons behind misreporting is an invaluable finding for policy makers and public health officials involved in policy creation and program planning. Knowing not only rates of overreporting but also who cares enough about their exercise identity to overreport may allow for more focused policy initiatives. While a better measure of exercise (and other normative behaviors) is certainly a goal of this research, the

overreporting of normative behavior is more than just a methodological puzzle to be solved for the sake of improving measurement.

Bibliography

- Adams, Swann Arp, Charles E. Matthews, Cara B. Ebbeling, Charity G. Moore, Joan E. Cunningham, Jeanette Fulton, and James R. Hebert. 2005. "The Effect of Social Desirability and Social Approval on Self-Reports of Physical Activity." *American Journal of Epidemiology* 161:389-98.
- Ainsworth, Barbara E., David R. Jacobs, and Arthur S. Leon. 1992. "Validity and reliability of self-reported physical activity status: the Lipid Research Clinics questionnaire." *Medicine and Science in Sports and Exercise* 25:92-98.
- Albanes, Demetrius, Joan M. Conway, Philip R. Taylor, Paul W. Moe, and Joseph Judd. 1990. "Validation and Comparison of Eight Physical Activity Questionnaires." *Epidemiology* 1:65-71.
- Aquilino, William S., and Leonard LoSciuto. 1990. "Effects of Interview Mode on Self-reported Drug Use." *Public Opinion Quarterly* 54:362-95.
- Baranowski, Tom. 1988. "Validity and Reliability of Self Report Measures of Physical Activity: An Information-Processing Perspective." *Research Quarterly for Exercise and Sport* 59:314-327.
- Bassett, David R. 2000. "Validity and Reliability Issues in Objective Monitoring of Physical Activity." *Research Quarterly for Exercise and Sport* 71:30-36.
- Belli, Robert F., Michael W. Traugott, Margaret Young, and Katherine A. McGonagle. 1999. "Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring." *Public Opinion Quarterly* 63:90-108.
- Bernstein, Robert, Anita Chadha, and Robert Montjoy. 2001. "Overreporting Voting: Why it happens and why it matters." *Public Opinion Quarterly* 65:22-44.
- Bolger, Niall, Angelina Davis, and Eshkol Rafaeli. 2003. "Diary Methods: Capturing Life as it is Lived." *Annual Review of Psychology* 54:579-616.
- Brenner, Philip S. 2011. "Identity Importance and the Overreporting of Religious Service Attendance: Multiple Imputation of Religious Attendance Using the American Time Use Study and the General Social Survey." *Journal for the Scientific Study of Religion* 50:103-115.
- Brenner, Philip S. and John D. DeLamater. 2012. "Using SMS Text Messaging To Collect Time Use Data." 67th Meeting of the American Association of Public Opinion Research, Orlando FL (May).

- Burke, Peter J. 1980. "The self: Measurement implications from a symbolic interactionist perspective." *Social Psychology Quarterly* 43:18-29.
- Cannell, Charles F., K. H. Marquis, and A. Laurent. 1977. A summary of studies of interviewing methodology. *Vital and Health Statistics, Series 2*:1-68.
- Cannell, Charles F., Peter V. Miller, and Lois Oksenberg. 1981. "Research on Interviewing Techniques." *Sociological Methodology* 11:389-437.
- Chase, David R. and Geoffrey C. Godbey. 1983. "The accuracy of self-reported participation rates." *Leisure Studies* 2:231-35.
- Chaves, Mark and Laura Stephens. 2003. "Church Attendance in the United States" in *Handbook of the Sociology of Religion*. Michele Dillon, ed. New York: Cambridge University Press.
- Durante, Richard and Barbara E. Ainsworth. 1996. "The recall of physical activity: using a cognitive model of the question-answering process." *Medicine and Science in Sports and Exercise* 28:1282-91.
- Ervin, Laurie H. and Sheldon Stryker. 2001. "Theorizing the Relationship between Self-Esteem and Identity." in *Extending Self-Esteem Theory and Research: Sociological and Psychological Currents*. Timothy J. Owens, Sheldon Stryker, and Norman Goodman, eds. New York: Cambridge University Press.
- Groves, Robert M. 1989/2004. *Survey Errors and Survey Costs*. New York: Wiley
- Hadaway, C. Kirk, Penny Long Marler, and Mark Chaves. 1998. "Overreporting Church Attendance in America: Evidence that Demands the Same Verdict." *American Sociological Review* 63: 122-30.
- Jacobs, David R., Barbara E. Ainsworth, Terryl J. Hartman, and Arthur S. Leon. 1992. "A simultaneous evaluation of 10 commonly used physical activity questionnaires." *Medicine and Science in Sports and Exercise*, 25:81-91.
- Klesges, Robert C., Linda H. Eck, Michael W. Mellon, William Fulliton, Grant W. Somes, and Cindy L. Hanson. 1990. "The accuracy of self-reports of physical activity." *Medicine and Science in Sports and Exercise*, 22:690-97.
- Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity." *Public Opinion Quarterly* 72:847-865.

- Leenders, Nicole Y. J. M., W. Michael Sherman, H. N. Nagaraja, and C. Lawrence Kien. 2001. "Evaluation of methods to access physical activity in free-living conditions." *Medicine and Science in Sports and Exercise*, 33:1233-40.
- Matthews, Charles E., Patty S. Freedson, James R. Hebert, Edward J. Stanek III, Philip A. Merriam, and Ira S. Ockene. 2000. "Comparing physical activity assessment methods in the Seasonal Variation of Blood Cholesterol Study." *Medicine and Science in Sports and Exercise*, 32:976-84.
- Matthews, Charles E. and Patty S. Freedson. 1995. "Field trial of a three-dimensional activity monitor: comparison with self-report." *Medicine and Science in Sports and Exercise*, 27:1071-78.
- Motl, Robert W., Edward McAuley, and Christine DiStefano. 2005. "Is social desirability associated with self-reported physical activity?" *Preventative Medicine* 40:735-39.
- Niemi, Iliris. 1993. "Systematic Error in Behavioral Measurement: Comparing Results from Interview and Time Budget Studies." *Social Indicators Research* 30:229-44.
- Patterson, Patricia. 2000. "Reliability, Validity, and Methodological Response to the Assessment of Physical Activity Via Self-Report." *Research Quarterly for Exercise and Sport* 71:15-20.
- Robinson, John P. 1999. "The Time-Diary Method: Structure and Uses." in *Time Use Research in the Social Sciences*. W.E. Pentland, A.S. Harvey, M.P. Lawton and M.A. McColl, eds. New York: Kluwer/Plenum.
- , 1985. "The Validity and Reliability of Diaries versus Alternative Time Use Measures." in *Time, Goods, and Well-Being*, F.T. Juster and F.P. Stafford, eds. Ann Arbor, MI: Institute for Social Research.
- Rzewnicki, Randy, Yves Vanden Auweele, and Ilse De Bourdeaudhuij. 2002. "Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample." *Public Health Nutrition* 6:299-305.
- Sallis, James F. and Brian E. Saelens. 2000. "Assessment of Physical Activity by Self-Reports: Status, Limitations, and Future Directions." *Research Quarterly for Exercise and Sport* 71:1-14.
- Schuman, Howard. 1982. "Artifacts are in the eye of the beholder." *American Sociologist*, 17:21-28
- Shephard, R. J. 2003. "Limits to the measurement of habitual physical activity by questionnaires." *British Journal of Sports Medicine* 37:197-206.

- Stinson, Linda L. 1999. "Measuring How People Spend Their Time: A Time Use Survey Design." *Monthly Labor Review* 122:12-19.
- Stocké, Volker. 2007. "Response Privacy and Elapsed Time Since Election Day as Determinants for Vote Overreporting." *International Journal of Public Opinion Research* 19:237-46.
- Stryker, Sheldon. 1980/2003. *Symbolic Interactionism: A Social Structural Version*. Caldwell, NJ: The Blackburn Press.
- Stryker, Sheldon and Peter J. Burke. 2000. "The Past, Present, and Future of an Identity Theory." *Social Psychology Quarterly* 63:284-97.
- Stryker, Sheldon and Richard T. Serpe. 1982. "Commitment, Identity Salience, and Role Behavior: Theory and Research Example." in *Personality, Roles, and Social Behavior*, William Ickes and Eric S. Knowles, eds. New York: Springer-Verlag.
- 1994. "Identity Salience and Psychological Centrality: Equivalent, Overlapping, or Complimentary Concepts?" *Social Psychology Quarterly*, 57:1.
- Tourangeau, Roger. 1984. "Cognitive Science and Survey Methods." *Cognitive Aspects of Survey Design: Building a Bridge Between Disciplines*, T. Jabine, M. Straf, J. Tanur, and R. Tourangeau, eds. Washington, DC: National Academy Press.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.
- U.S. Department of Health and Human Services. 2008. *2008 Physical Activity Guidelines for Americans*. Office of Disease Prevention and Health Promotion, ODPHP U0036.
- Washburn, Richard, Gregory W. Heath, and Allen W. Jackson. 2000. "Reliability and Validity Issues Concerning Large-scale Surveillance of Physical Activity." *Research Quarterly for Exercise and Sport* 71:104-113.
- Williams, Ellen, Robert C. Klesges, Cindy L. Hanson, and Linda Eck. 1989. "A Propective Study of the Reliability and Convergent Validity of Three Physical Activity Measures in a Field Research Trial." *Journal of Clinical Epidemiology* 42:1161-70.
- Weisberg, Herbert F. 2005. *The Total Survey Error Approach : A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press.
- Woodberry, Robert D. 1998. "When Surveys Lie and People Tell the Truth: How Surveys Oversample Church Attenders." *American Sociological Review* 63: 119-21.
- Zerubavel, Eviatar. 1989. *The Seven Day Circle: The History and Meaning of the Week*. Chicago: University of Chicago Press.

Zuzanek, Jiri and Bryan J. A. Smale. 1999. "Life-Cycle and Across-the-Week Allocation of Time to Daily Activities." in *Time Use Research in the Social Sciences*. W.E. Pentland, A.S. Harvey, M.P. Lawton and M.A. McColl, eds. New York: Kluwer/Plenum.

Table 1.
Number and percentage of respondents, by subtype

	Web		Text	
Validated exercisers	20	(22%)	22	(33%)
Overreporters	43	(48%)	5	(7%)
Admitted non-exercisers	22	(25%)	33	(49%)
Underreporters	4	(4%)	7	(10%)
	89	(100%)	67	(100%)

Table 2.
Descriptive statistics for Web respondents, by subtype

	N	Mean	s.d.	Median
Validated exercisers	20	4.4	0.68	4.5
Overreporters, all	43	4.2	0.81	4.0
Nonexercising overreporters	24	4.4	0.71	4.5
Admitted non-exercisers	22	3.8	0.92	4.0
Underreporters	4	3.0	1.41	3.5

Figure 1.

Percentage of overreporters and validated exercisers claiming levels of identity importance

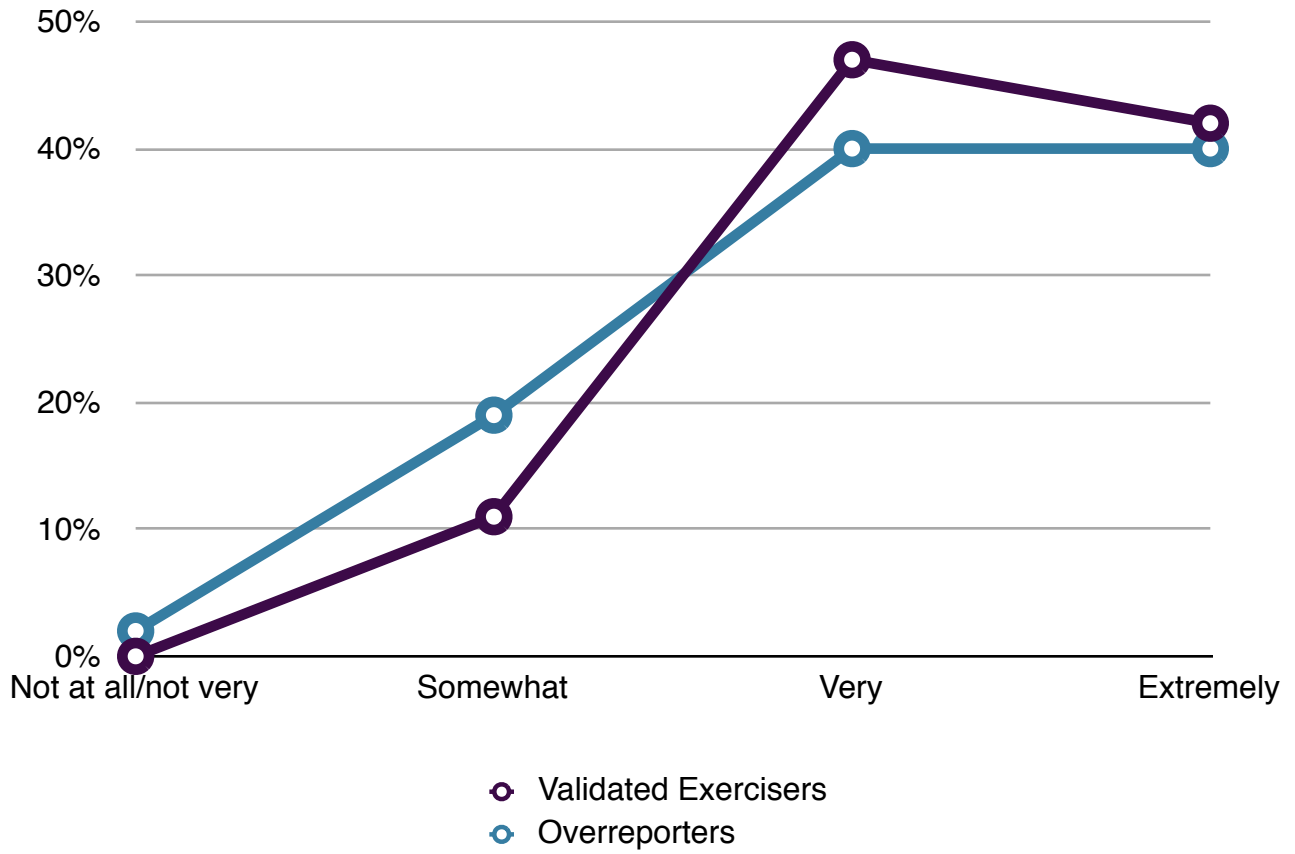


Figure 2.

Percentage of overreporters and admitted non-exercisers claiming levels of identity importance

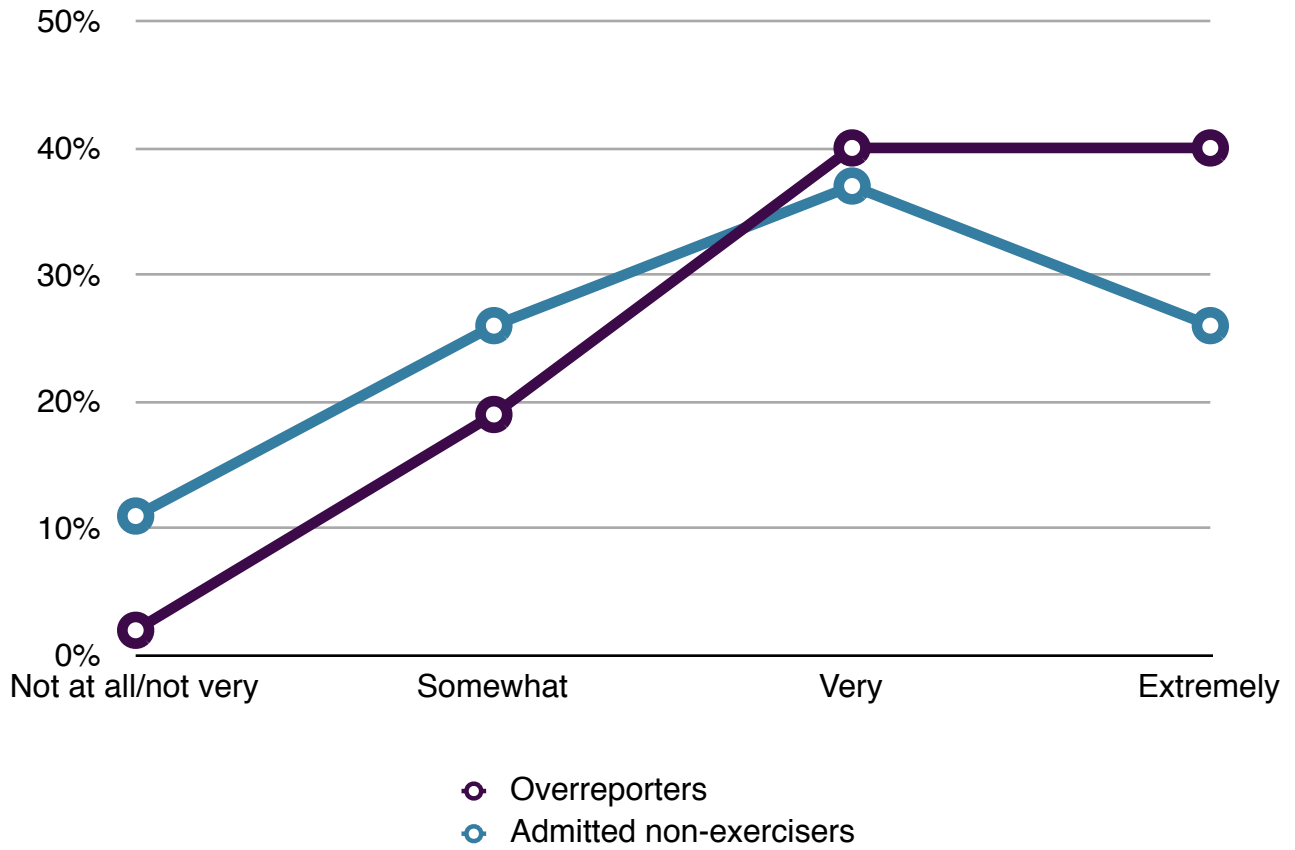
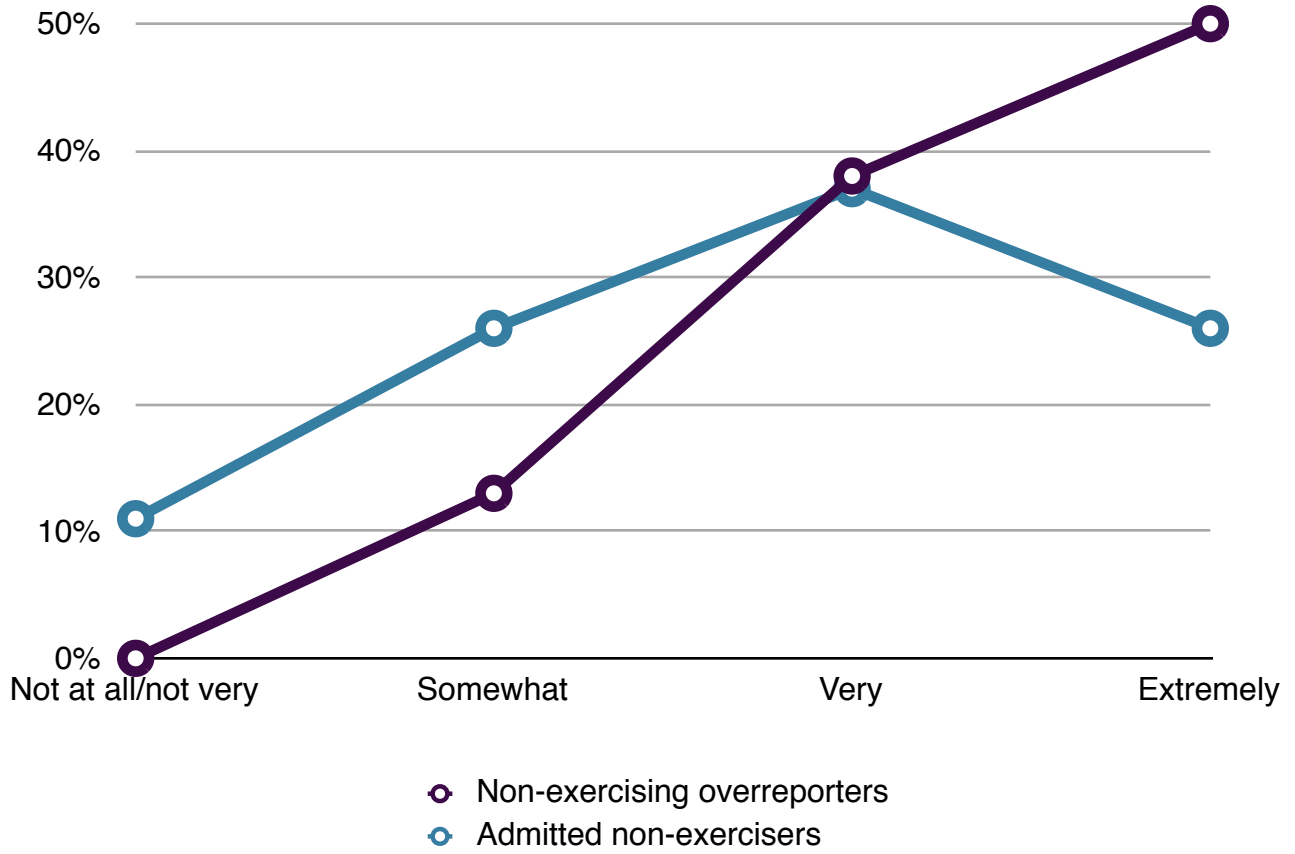


Figure 3.

Percentage of non-exercising overreporters and admitted non-exercisers claiming levels of identity importance



Center for Demography and Ecology
University of Wisconsin
1180 Observatory Drive Rm. 4412
Madison, WI 53706-1393
U.S.A.
608/262-2182
FAX 608/262-8400
comments to: delamater@ssc.wisc.edu
requests to: cdepubs@ssc.wisc.edu