

Review

The validity of self-report measures of proenvironmental behavior: A meta-analytic review



Christine Kormos*, Robert Gifford

Department of Psychology, University of Victoria, Canada

ARTICLE INFO

Article history:

Available online 10 October 2014

Keywords:

Meta-analysis
Proenvironmental
Self-report
Objective
Observed
Behavior

ABSTRACT

Do self-reports match objective behavior? We performed a meta-analysis to quantify the association between self-reported and objective measures of proenvironmental behavior, and to evaluate the moderating influence of two socio-demographic and seven methodological moderators. Data from 6260 individuals or households, involving 19 measures of association in 15 studies, revealed a positive and nominally large (Cohen, 1988) effect size ($r = .46$). However, this means that 79% of the variance in the association between self-reported and objective behavior remains unexplained, which is especially troubling given the environmental context. We conclude that although this effect size is conventionally large, it is functionally small for testing theory and devising intervention campaigns, possibly leading researchers to draw misleading conclusions about the usefulness of theories that employ self-reports to predict objective behavior. These findings highlight a crucial need for research that strengthens the validity of self-reports for well-defined types of environmental behavior.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

As environmental concern mounts, interest in understanding the psychological determinants of proenvironmental behavior has increased. Environmental psychologists and others in related disciplines have endeavored to identify the most influential predictors of proenvironmental behavior and to develop models that represent the role of intrapersonal (e.g., attitudes, beliefs, and values), interpersonal (e.g., social norms), and external factors (e.g., rewards and punishments) (e.g., Gifford, 2006). Self-reports often are employed as measures of proenvironmental behavior (see Steg & Vlek, 2009 for review). They usually are obtained through surveys, questionnaires, or interviews, and they assess participants' subjective estimates of their own behavior. Of course, the use of self-reports rests on the assumption that they accurately reflect individuals' *actual* behavior – an assumption that has received mixed empirical support (e.g., Hamilton, 1985; Warriner, McDougall, & Claxton, 1984). Some studies suggest that self-reports are adequate predictors of actual behavior (e.g., Corral-Verdugo & Figueredo, 1999), whereas others suggest the contrary (e.g., Fuj, Hennessy, & Mak, 1985).

Thus, construct validity is a key concern with the use of self-reports, given that researchers strive to employ the best measures of proenvironmental behavior. *Construct validity* refers to the degree to which a measure (e.g., self-reported water consumption) accurately assesses the intended construct (e.g., actual water consumption). The establishment of sound construct validity is an essential requirement for any rigorous line of research – in fact, without it the chosen variable simply cannot be said to represent the construct that it purports to measure.

The various appeals of self-report measures (primarily their ease of use) make it unlikely that their usage in the literature will decrease anytime soon, and therefore evaluating and understanding the nature and degree of error that they introduce is crucial. For this reason, this meta-analysis assesses the association between self-reported and objective measures of proenvironmental behavior in studies that have employed both types of measures to evaluate the same behavior. Additionally, it investigates several socio-demographic and methodological characteristics that may magnify or reduce the error in self-report measures.

1.1. Advantages of self-report measures

Self-reports are the preferred method of data collection for the majority of researchers, owing to their low cost, relative ease of use, and flexibility. Indeed, simply asking participants to report, for example, how often they engage in a particular environmentally

* Corresponding author. Department of Psychology, P. O. Box 3050, University of Victoria, Victoria, British Columbia, V8W 3P5, Canada. Tel.: +1 250 472 4876; fax: +1 250 721 8929.

E-mail address: ckormos@uvic.ca (C. Kormos).

relevant behavior along a scale from 'Never' to 'Always' (e.g., De Young, 1990; Margai, 1997) is an easy way to obtain information about that behavior. Self-reports also allow researchers to investigate behaviors that may not otherwise be observable (see Tarrant & Cordell, 1997). Given these advantages, researchers often use self-report measures to ask participants to express the frequency with which they engage in a list of proenvironmental behaviors (e.g., Gatersleben, Steg, & Vlek, 2002; Kaiser, Frick, & Stoll-Kleemann, 2001).

1.2. Disadvantages of self-report measures

The broader survey methodology literature suggests that self-reports are only weakly associated with actual behavior (e.g., Peterson & Kerin, 1981). Inaccuracies may stem from a variety of sources. For example, self-report measures may be prone to exaggeration. Some evidence suggests that individuals tend to over-report their proenvironmental behavior (Barr, 2007; Fuj et al., 1985; Geller, 1981; Warriner et al., 1984), and social desirability bias has been suggested as a cause for this over-reporting and thus an important limitation of self-report measures of proenvironmental behavior as compared to objective measures (Stern & Oskamp, 1987; Tarrant & Cordell, 1997). Socially desirable responding is defined as "the tendency of subjects to attribute to themselves in self-description, personality statements with socially desirable scale values, and to reject those with socially undesirable scale values" (Edwards, 1957, p. vi). Recent research, however, has provided explicit evidence that social desirability appears to have a minimal and non-significant effect on self-reported proenvironmental behavior and no moderating effect on the relationship between self-reported environmental attitudes and proenvironmental behavior (Milfont, 2009).

Another disadvantage of self-report measures is that they are subjective by nature; descriptors such as 'Often,' may mean different things to different participants. In addition, survey instruments are best designed to assess attitudes and beliefs, and therefore even when respondents are explicitly asked to report their behavior, attitudes often seep into their responses. Some have also suggested that self-reports may largely reflect individuals' perceptions of their behavior (Olson, 1981), behavioral intentions (Lee, 1993), or other – sometimes false – beliefs and attitudes (Rathje, 1989), rather than objective behavior. Finally, limited memory or knowledge may also reduce the accuracy of self-reports (e.g., see Warriner, McDougal, & Claxton, 1984).

1.3. Self-report validity as an issue of concern in other disciplines

Given that the advantages of self-report measures make them common research tools in these other areas as well, these other disciplines must also grapple with similar issues pertaining to their use. As such, the validity of self-report data is a question which has received considerable research attention in a variety of other disciplines, including public opinion polling, consumer research, and – perhaps most notably – health research. This broader consideration of self-report validity warrants some discussion here.

In studying health-related behaviors, for example, self-reports are frequently used to privately assess what may be sensitive or personal issues. For example, health researchers have explored the validity of self-reports for assessing a variety of behaviors, including sexual behavior (e.g., Brener, Billy, & Grady, 2003). Research examining health-risk behaviors (e.g., unprotected sex and drug use) in adolescents often relies, by necessity, on self-report questionnaires. However, Brener et al. (2003) asserted that both the accuracy and truthfulness of such data might be compromised when these behaviors are measured through self-report, given that

situational factors, such as a lack of privacy when responding to the questionnaire, may cause respondents to provide inaccurate information, or that respondents may inaccurately recall information, or may over- or under-report behaviors due to social desirability or the desire to avoid providing sensitive information. As an attempt to improve self-report accuracy, Denny et al. (2008) conducted a national survey of health and wellbeing among youth in which adolescents were provided with individual computer tablets. Students who preferred the tablet over a standard lab computer reported feeling more privacy and finding it easier to answer questions truthfully, thus potentially providing more valid self-report data (Denny et al., 2008).

For other health-related behaviors, such as physical activity, Affuso et al. (2011) found large discrepancies between self-reported physical activity levels and objectively measured activity via accelerometry. Thus, Affuso et al. concluded that self-reports are not a valid measure of exercise levels given individuals' tendency to under-report sedentary behaviors and that objective measures should be used instead. In the case of smoking behavior, however, Studts et al. (2006) demonstrated that self-reported smoking habits served as a valid measurement of actual smoking behavior, as evaluated via objective testing of urine samples, a finding which may surprise some given that self-reporting is often questioned due to social desirability and medical disapproval, which is thought to result in under-reported smoking behaviors. Thus, health researchers also struggle with issues related to whether subjective measures accurately capture their targets behaviors.

The validity of self-report measures is also of concern in the discipline of marketing and business research, where the accuracy of such measures can provide valuable information on consumers' brand preference, purchasing behavior, and product usage. Nencyz-Thiel, Beal, Ludwichowska, and Romaniuk (2013), for example, examined the accuracy of self-reported television viewing as compared to objectively measured minute-per-minute data on actual programs watched. Their research concluded that discrepancies exist between the objective measures and individuals' reporting of the amount of television viewing as well as the specific programs viewed.

Therefore, researchers across a variety of disciplines outside of environmental psychology must address similar issues pertaining to self-report validity and, as such, have undertaken efforts to validate their self-report measures and explored means to minimize the discrepancy between self-reports and objective measures of behavior. Within environmental psychology research outside of the domain of proenvironmental behavior, researchers have found that judgments of environmental features do not always correspond accurately with objective measures of the same features (McCormack, Cerin, Leslie, du Toit, & Owen, 2008). Those researchers suggest that future research should include measures of perceived and objective environmental attributes.

1.4. Description of relevant studies

Therefore, similar to the fields of health and marketing, environmental psychology researchers rely heavily on self-reporting, which can be validated in different ways. Several types of unobtrusive measures have been used to compare self-reports with more objective measures and these can be broadly classified into three categories: device measurements, observations made by trained observers, and peer ratings.

1.4.1. Device measurements

Energy use and water consumption are two behavioral domains that lend themselves well to the evaluation of the accuracy of self-reports through device measurements – specifically, via meter

readings. Many studies conducted in the 1980s used this method of data collection. For example, self-reported household energy consumption (for electricity, oil, and gas) was strongly correlated with utility company data (Warriner et al., 1984). However, Fuj et al. (1985) observed that participants tended to over-report their electricity conservation efforts, when compared to meter readings. Furthermore, the correlation between stated change to household electricity consumption during the last year and the observed change in electricity usage was quite low ($r = .06$). Consistent with this, self-reported water consumption has been shown to be only weakly correlated with the change in actual water consumption, $\chi^2_{(6)} = 1.9, p = .01$ (Hamilton, 1985). Other such measurements may be obtained by a variety of devices, including GPS readings of transportation behavior (Bolderijk, Knockaert, Steg, & Verhoef, 2011) and satellite readings of deforested land (Vadez et al., 2003).

1.4.2. Trained observers

Other studies have obtained objective measures through direct observations made by trained individuals of participants' behaviors. For example, self-reported recycling and the observed frequency of this behavior was weakly associated ($r = .14$; Corral-Verdugo, 1997), as was self-reported recycling with the amount of material put out for recycling ($r = .16$; McGuire, 1984). On the other hand, others have found a much higher correlation between self-reported and in-home observations of the re-use of glass, clothing, and metal (mean $r = .63$; Corral-Verdugo & Figueredo, 1999).

1.4.3. Peer ratings

More recently, some researchers have explored the accuracy of self-reports using ratings by people who are close to the participant (e.g., spouses and housemates). Peer ratings differ from the above in that they are made by an individual who is *not* affiliated with a research team. This approach is particularly well-suited to pro-environmental behavior given that many relevant behaviors are performed in private. For example, individuals' self-reports of four environmental behaviors were strongly associated in one study with a report by their spouses ($r = .45$; Lam & Cheng, 2002). However, the frequency of self-reported proenvironmental behavior can be considerably higher than peer-reported behavior (Chao & Lam, 2011).

1.5. Implications for theory development

The accuracy of self-reports as measures of proenvironmental behavior is especially important given that they are often used as the key criterion variable in research and theory development; for example, the theory of planned behavior (Ajzen, 1985, 2005), one of the most widely used models for predicting proenvironmental behavior, often uses self-reports as the dependent variable and assumes that behavioral intention is the closest and most direct psychological determinant of behavior, and that intention is, in turn, causally determined by three factors: attitude towards the behavior (determined by values and beliefs), social norms, and perceived behavioral control. The theory of planned behavior has been applied to predict a variety of environmental behaviors, such as recycling (e.g., Boldero, 1995) and public transportation use (e.g., Heath & Gifford, 2002).

In a meta-analysis that examined the efficacy of the theory of planned behavior (Armitage & Conner, 2001) – in which the majority of the 185 studies included were health-related but nine had proenvironmental behavior as the outcome measure – the theory of planned behavior explained 10% more variance in self-reported behavior than it explained in objective or observed behavior ($R^2 = .31$ versus $R^2 = .21$). Consistent with this finding, a recent study found that the theory of planned behavior predicted self-

reported proenvironmental behavior considerably better than it predicted peer-reported behavior ($R^2 = .81$ versus $R^2 = .10$; Chao & Lam, 2011). Thus, self-report validity can have implications for theory development, given that prediction models may inadvertently exaggerate the usefulness of a theory for explaining objective proenvironmental behavior when the criterion variable is a self-report of proenvironmental behavior. Of course, methodological issues, such as content overlap (i.e., where the theory of planned behavior measures and self-report measures share similar wording) and similarities in response format across measures can also affect shared variance. For example, the mono-method bias, which is shared variance among constructs that are all measured in the same way, has been demonstrated in the environmental psychology literature (Evans, 1999). Factors such as these can methodologically inflate the correlations between measures and thus could account for the differences in the proportion of variance explained by the theory of planned behavior.

1.6. Potential moderators

Given the possible limitations of self-reports, understanding the factors that may influence their validity is important. Two socio-demographic variables may moderate the association between self-reported and objective behavioral measures.¹ First, participant gender may influence validity because previous studies have demonstrated gender differences in proenvironmental behavior, in that females typically report engaging in more proenvironmental behaviors than males (e.g., Lam & Cheng, 2002; Zelezny, Chua, & Aldrich, 2000). Furthermore, health research has demonstrated differences in self-report validity between male and female participants; for instance, when asked to provide bodily measurements, women tend to under-report their weight more often than men, and men tend to over-report their height more often than women (Elgar & Stewart, 2008). Second, participant age should be evaluated because several studies have shown that younger people tend to report being more environmentally concerned than older people (e.g., Van Liere & Dunlap, 1980), although some studies have found no actual age difference in certain types of proenvironmental behavior, such as household electricity consumption (e.g., Fuj et al., 1985). But, again, evidence from the domain of health research has shown that age influences self-report validity; for example, older individuals are more likely to over-report their height, as compared to younger individuals (Elgar & Stewart). Age-related discrepancies in self-reported versus objective measures are also observed in self-reported vaccination records as compared to electronic medical records, such that adults over age 50 are slightly less likely to underreport vaccinations than adults under age 50 (Rolnick et al., 2013).

In addition, seven methodological variables should be considered. First, the number of items in the self-report measure may moderate the association. For example, some studies that have used a large number of items, such as Kaiser, Frick, and Stoll-Kleemann's (2001) 65-item measure, have reported strong correlations (i.e., $r = .81$), whereas other studies that used only one item have reported much weaker correlations ($r = .06$, Fuj et al., 1985; $r = .18$, Hamilton, 1985). Second, the number of response options may moderate the self-report-proenvironmental behavior relation. Self-reports have been particularly accurate at predicting pro-environmental behavior when they are dichotomized, such as '*I do*'

¹ Some research has shown that report accuracy is greater among high socioeconomic status (SES) participants than low SES participants (e.g., Hamilton, 1985), but unfortunately there was not enough data present in the sample studies to evaluate this variable.

versus 'I don't' (e.g., Kaiser, Doka, Hofstetter, & Ranney, 2003), whereas some studies that use non-dichotomized scales, such as Likert-style scales, with a range of response options have yielded weaker correlations (e.g., Corral-Verdugo, 1995). Third, whether the unit of analysis is the individual or a household may moderate the association given that it may be more difficult for people to estimate others' behavior as opposed to their own.

Several additional exploratory moderators are important to consider. Fourth, details about the nature of the objective measure (i.e., whether it is obtained by a mechanical device, trained observers, or peer ratings) may explain some variance among observed effect sizes. Fifth, the type of proenvironmental behavior (e.g., recycling, energy usage, water usage, or transportation) may influence the degree of congruence between the two types of measures. For instance, self-reports have been shown to be either greater or lesser than peer ratings depending on the type of proenvironmental behavior examined (e.g., Lam & Cheng, 2002). Even within one behavioral domain (i.e., waste reduction), the association between self-reported and observed re-use and recycling behavior has differed ($r = .08$ and $r = .25$, respectively; Corral-Verdugo, Bernache, Encinas, & Garibaldi, 1995). Furthermore, inconsistencies in the literature also suggest that the type of behavior examined may moderate the degree of observed congruence between these two types of measures. Sixth, whether or not the validity of self-reports changed over the time period of the studies included in the sample is worth exploring. Thus, the year of study publication was included as an exploratory moderator to assess whether the validity of self-reports has changed over the time period of the studies included in the sample. Research in the health research domain suggests that date of study publication is an important variable to assess; for example, a study assessed progress in the reliability and validity of self-report measures of HIV-related sexual behavior since 1990, in an effort to make recommendations for research and practice (Weinhardt, Forsyth, Carey, Jaworski, & Durant, 1998). Last, study location (i.e., country), may also be useful to consider given the possibility that self-report validity may vary across countries.

1.7. Aim of the current study

This meta-analysis evaluated all studies that could be located which included a self-reported and objective measure of the same proenvironmental behavior – both assessed using the same participants – to examine the overall association between the two types of measures. In doing so, it investigated the degree to which self-reports reflect objective behavior, as well as whether or not a systematic tendency exists for self-reports to under- or overestimate objective behavior. An additional objective was to assess the potential influence of two socio-demographic and seven methodological moderators.

2. Method

2.1. Selection of studies

This meta-analysis included only studies that examined associations between a self-reported measure of an ecologically relevant behavior and an objective measure of the same behavior. For the purposes of this analysis, ecologically relevant behavior encompasses any action relevant to environmental and resource sustainability, either in terms of behaviors within the household or within the broader community. Self-report measures ranged, for example, from individuals' assessments of their water usage, energy consumption, and recycling behavior to individuals' estimates of land deforestation. These measures were required to be reports of participants' past or present behavior, rather than their future

behavioral intentions.² Objective measures included, for example, device measurements (e.g., meter readings of electricity and water usage), observations made by trained observers (e.g., in-home observations of re-use or recycling behavior), and peer ratings (e.g., made by spouses or housemates).³

To be eligible for inclusion, several criteria had to be met.⁴ First, the self-report and the objective measure must have been taken from (or of) the same individual or household. Thus, studies in which these two measures were taken from different individuals or households were ineligible, as were those that used an aggregate objective measure (e.g., a population statistic or a measure for an entire block). Second, self-report and objective measure measures had to assess the same proenvironmental behavior. Third, articles had to be accessible in English, unless the appropriate information could be obtained through email from the authors. Fourth, studies had to provide primary, quantitative data, containing the information necessary to calculate the effect size for the association between the two types of measures. When an article met all these criteria except the last one, authors were emailed in an attempt to procure this information.

Relevant published and unpublished studies were identified through three means. First, online searches were performed in two databases (PsycINFO and ProQuest Dissertations and Theses) through March, 2011. Key truncated search terms (taking into account multiple spellings) included *self-report*, *stated*, *survey*, and *questionnaire* with *actual*, *observed*, *other-reported*, *measured* and *objective* with *behavior* and *response* with *environment*, *energy*, *water*, *recycle*, *electricity*, *household*, *proenvironmental*, *consumer*, *conserve*, *consumption*, *utility*, *sustainable*, *climate change*, *global warming*, *resource*, and *action*. These terms were searched in the Abstract field. In PsycINFO, the classification category (CC) number '4070' was entered to specify environmental issues and attitudes. In ProQuest, the truncated word *environment* was searched as a keyword index term. This strategy generated 499 articles to be evaluated (197 from PsycINFO and 302 from ProQuest). Two members of the research team screened all article titles and abstracts to determine whether or not they met the basic eligibility criteria. If insufficient information was available in the abstract to determine eligibility, the full-text article was retrieved and evaluated. Second, a request for relevant published and unpublished articles was made on a popular subject listserve (i.e., APA Division 34: Population and Environmental Psychology). This appeal yielded two articles that were subsequently included in the analysis. Third, reference lists from relevant articles as well as those from review articles were checked for studies that may have been inadvertently missed.

Following some initial exclusions, based on the eligibility criteria, the pool of identified studies was reduced to 106 potentially useful articles. Full-text versions of these articles were subsequently evaluated more thoroughly and further exclusions were made as necessary. Of the 106 articles, 15 studies with data from 6260 individuals or households were included in the final analysis (see Table 1). Three of these studies contained multiple distinct behavioral measures, and therefore a total of 19 association values were examined in the analyses to follow.

² Behavioral intention sometimes is used as a measure of objective behavior (e.g., Stern, Dietz, Kalof, & Guagnano, 1995), and yet prudence dictates caution given that it is currently unclear exactly how self-reported behavior and behavioral intention are related.

³ Articles that ostensibly gathered an objective measure of behavior by asking participants to provide a self-report of their "actual behavior" were ineligible.

⁴ Not all relevant studies could be included in the analysis for reasons related to the inclusion criteria. For instance, several classic studies such as Geller (1981) and Luyben (1982) were ineligible.

2.2. Data extraction

The first author coded all studies for the statistics needed to compute effect sizes as well as the moderation analyses. For the moderation analyses, two participant characteristic variables were coded for each study: percent of male participants and mean age of participants. Additionally, seven methodological characteristics were coded for each study: the number of items in the self-report measure, number of response options, whether the participants were individuals or households, the nature of the objective measure (device measurement, trained observers, and peer ratings), type of behavior (e.g., recycling and energy usage), year of publication, and study location. The coding and data entry were conducted in an Excel spreadsheet and later imported into the Comprehensive Meta-Analysis Version 2 software program (Borenstein, Hedges, Higgins, & Rothstein, 2005).

2.3. Calculating effect sizes

Effect sizes were calculated for each study and for the overall effect size using Comprehensive Meta-Analysis. For the majority of studies, Pearson's r values for the correlation between the self-report and objective measure were extracted from the article and then underwent a Fisher's z -transformation. In the three studies that reported on a few distinct types of proenvironmental behavior, the values were treated separately.⁵ Furthermore, an odds ratio was calculated for one study, and a chi-square value was obtained from another study. Both of these values were then entered into Comprehensive Meta-Analysis and used to calculate individual effect sizes and the overall effect size, both represented as Pearson's r .

2.4. Testing for systematic bias

To evaluate the nature and degree of potential bias in self-report measures (i.e., the extent to which self-reports tend to under-report, over-report, or accurately report objective measures), each dataset in the meta-analysis was individually examined. Considerable variability in study design and data presentation made it impossible to systematically assess all studies in a standardized manner (Table 1). In particular, four studies were excluded from this analysis because their self-report measure could not be meaningfully compared to the objective measure (e.g., in the case of a Likert-style scale for the self-report and a continuous objective measure), and four other studies were excluded because of insufficient data (mean values were not included in the articles and authors could not be reached to supply them).

Of the remaining 11 studies, four presented cross-tabulation matrices of agreement between self-reports and objective measures; in these cases, the number of participants who over-reported, under-reported, and accurately reported their behavior were summed, and that total was divided by the total number of participants to yield the percentage of over-, under-, or accurate reporting in the sample. In the last seven studies, the self-report and objective measures were presented using a common metric (e.g., the same Likert scale or frequency of re-use items), which allowed them to be directly compared in a ratio to determine the degree of under- or over-reporting.⁶

⁵ When a study included bivariate regressions for multiple related items, the values were averaged to obtain an overall correlation value.

⁶ The Vadez et al. (2003) study was not included in this analysis because its measures assessed deforestation, which is in the opposite direction of pro-environmental behavior.

3. Results

3.1. Descriptive statistics

In total, 6260 individuals or households participated in the 15 studies included in the meta-analysis. Table 2 displays the descriptive statistics for each study. On average, participants were 38.2 years of age (range 24.2–45.2 years, $k = 6$ studies or cases⁷) and 31.49% of the participants were male (range 0–100%, $k = 9$). Year of study publication ranged from 1984 to 2011. The mean number of items in the self-report measure was 7.95 (range = 1–65, $k = 19$), and the mean number of response options was 4.5 (range = 2–7, $k = 12$). Forty-seven percent ($k = 9$) of the outcome measures were made by trained observers (e.g., in-home observations of re-use and recycled products), 42% ($k = 8$) were made by device measurement (e.g., meter reading or GPS reading), and 11% ($k = 2$) were peer ratings (i.e., by a housemate or spouse).

In terms of the behaviors under study, 32% ($k = 6$) of outcome measures were of waste-relevant behavior (i.e., re-use or recycling), 26% ($k = 5$) were of energy usage (i.e., electricity, gas, or oil), 11% ($k = 2$) were of water usage, 5% ($k = 1$) each were of food consumption, deforestation, and transportation, and 16% ($k = 3$) were measures comprised of a mixture of proenvironmental behaviors.

3.2. Overall effect size and heterogeneity

Using the random-effects model, the average point estimate was found to be $r = .46$, indicating a large effect size (Cohen, 1992). Self-reported measures of proenvironmental behavior were, overall, strongly correlated with objective measures (95% CI = .28–.60, Z -value = 4.66, $p < .001$; Fig. 1).⁸ The coefficient of determination ($r^2 = .21$) indicates that 21% of the variance in self-reports was linearly associated with variance in objective measures.

Significant heterogeneity was detected among studies, $Q(18) = 1273.57$, $p < .001$, based on a measure of weighted squared deviations. Furthermore, $I^2(98.59)$ indicated that 99% of the observed variance was explained by true systematic effect size differences (as opposed to error) between studies. Both statistics suggest the need to explore possible moderating variables.

3.3. Moderator analyses

The percent of male participants in a study emerged as a significant moderator, $Q_{\text{model}}(1) = 7.13$, $p = 0.01$; studies with more males had stronger self-report-objective measure correlations than those with fewer males.⁹ Average participant age, however, did not significantly moderate the association under study, $Q_{\text{model}}(1) = .29$, ns .

The number of items in the self-report measure, $Q_{\text{model}}(1) = 1.21$, ns , the number of response options, $Q_{\text{model}}(1) = 1.01$, ns , and whether participants were individuals or households, $Q_{\text{model}}(1) = 0.58$, ns , did not account for significant variation across studies. Additionally, the nature of the objective measure (device measurement, trained observer, and peer rating) was not a significant moderator, $Q_{\text{value}}(2) = 0.60$, ns .¹⁰

⁷ Data on participant age were available for six of the 15 studies.

⁸ As shown in Table 2, Corral-Verdugo (1997) and Corral-Verdugo (1995) used the same participants for each of their assessments, and therefore 150 participants were double-counted in this analysis.

⁹ This moderation is still significant without the inclusion of the Vadez et al. (2003) study on deforestation, $Q_{\text{model}}(1) = 7.25$, $p = .01$.

¹⁰ This moderation is still not significant without the inclusion of the Vadez et al. (2003) study on deforestation, $Q_{\text{value}}(2) = .29$, ns .

Table 1Study design details and methods of data collection for each dataset included in the meta-analysis ($N = 19$).

Source	Behavior	Self-report measure	Number of items in self-report	Number of response options	Objective measure	Nature of objective measure	Accuracy of PEB reporting
Cross-tabulation (number of participants)							
Fuj et al. (1985)	Energy usage (Electricity)	Change over past year	1	2 “Yes” or “No”	Change over past year from meter reading	Device measurement	Accurate (53%) Over (30%) Under (17%)
Hamilton (1985)	Water usage	Change over past year	1	4 “Increased,” “Decreased,” or “Stayed the same,” or “Don’t know”	Change over past year from meter reading	Device measurement	Accurate (46%) Over (28%) Under (26%)
Kaiser et al. (2001)	Varied ^a	14 PEBs	65	Mixed ^b	Various PEBs and items observed	Trained observer	Accurate (89%) Over (5%) Under (6%)
Cross-tabulation (number of retrofit measures installed)							
Hirst and Goeltz (1985)	Energy usage (Retrofit measures)	The number of retrofit measures installed	9	n/a	In-home observation by utility company of number of retrofit measures installed	Trained observer	Accurate (81%) Over (9%) Under (10%)
Common metric							
Chao and Lam (2011)	Varied	Five types of PEB	5	7 “Never” to “Always”	Same as SR but other-reported (roommate)	Other-report	Over (132%) [Mean SR (5.22)/mean other-report (3.96)]
Corral-Verdugo and Figueroa (1999)	Waste (Re-use)	Quantity of reuse of three materials	4	4	In-home observation of number of reused objects in household	Trained observer	Under (91%) [Mean SR (4.87)/mean observed reused items (5.37)]
Lam and Cheng (2002)	Varied ^c	Frequency of four PEBs	11	7 “Never” to “Very often”	Same as SR, but other-reported (spouse)	Other-report	Accurate (101%) [Mean SR (3.56)/mean other-report (3.54)]
Vadez et al. (2003)	Estimate of deforestation	Estimated area of forest cleared that year by head of household and plot owner.	1	—	Surface area and perimeter (acreage) of plots of land cleared	Device measurement	Over (88.5%) ^d [Mean SR (13.6 ha)/mean objective area cleared (15.36 ha)]
Warriner et al. (1984)	Energy usage (electricity)	Money spent on electricity in past year	1	—	Amount spent on electricity	Device measurement	Under [Discrepancy between measures = \$12.02]
Warriner et al. (1984)	Energy usage (gas)	Money spent on gas in past year	1	—	Amount spent on gas	Device measurement	Under [Discrepancy between measures = \$16.59]
Warriner et al. (1984)	Energy usage (oil)	Money spent on oil in past year	1	—	Amount spent on oil	Device measurement	Over [Discrepancy between measure = \$-22.3]
Unknown							
Bolderdijk, Knockaert, Steg, and Verhoef (2011)	Transport (speeding)	Driver Behavior Questionnaire related to speeding	3	6 “Never” to “Almost always”	GPS-based percent of distance driven at 5% or more above speed limit	Device measurement	Cannot compare
Corral-Verdugo (1995)	Waste (Re-use)	Reuse frequency of four materials	4	4 “Never” to “Always”	Number of reuse items observed through refuse analysis	Trained observer	Cannot compare
Corral-Verdugo (1995)			4				

Table 1 (continued)

Source	Behavior	Self-report measure	Number of items in self-report	Number of response options	Objective measure	Nature of objective measure	Accuracy of PEB reporting
Corral-Verdugo (1997)	Waste (Recycling)	Recycling frequency of four materials	8	4 "Never" to "Always"	Frequency of recycled items observed through refuse analysis	Trained observer	Cannot compare
	Waste (Reuse)	Frequency of re-use of eight materials		4 "Never" to "Always"	In-home observation of number of to-re-use items in household	Trained observer	Cannot compare
Corral-Verdugo (1997)	Waste (Recycling)	Frequency of recycling of six materials	8	4 "Never" to "Always"	In-home observation of number of to-recycle items in household	Trained observer	Cannot compare
Corral-Verdugo (2003)	Water usage	Observation of five water usage activities made by housewives	5	—	Meter reading	Device measurement	Insufficient data
Cote (1984)	Food consumption	Stated volume of 15 food items consumed by household	15	—	Observed through refuse analysis	Trained observer	Insufficient data
McGuire (1984)	Waste (Recycling)	Frequency of recycling aluminum and newspapers	4	6 "All" to "None"	Observed through refuse analysis	Trained observer	Insufficient data

^a Grouped into domains of power conservation, PEB consumer behavior, garbage inhibition, and ecological automobile use.

^b A yes/no format was used ($n = 30$) or responses were re-coded, collapsing "Never", "Seldom", and "Occasionally" to "No" and turning "Often" and "Always" to "Yes" responses.

^c Assessed consumption, recycling, information seeking, and political actions.

^d The SR is 88.5% of the OB, and so participants over-reported their land conservation.

The type of behavior under study approached significance as a moderating effect, $Q_{value(6)} = 12.26$, $p = .06$.¹¹ The validity of self-reports was greatest for deforestation-related behavior ($r = .73$), followed by energy usage ($r = .61$), varied behaviors ($r = .55$), food ($r = .31$), transportation ($r = .30$), water usage ($r = .29$), and waste behavior ($r = .28$). Year of study publication did not significantly moderate the association, $Q_{model(1)} = 0.04$, ns . Finally, the country of study location significantly moderated the association between self-reports and objective measures, $Q_{value(6)} = 32.18$, $p < .001$.¹² Across the 15 studies included in the analysis, five studies are from Latin America (namely, Mexico and Bolivia). Comparing the effect sizes between these two categories of countries reveals that the average of the 12 association values from non-Latin American countries ($r = .45$) is higher than the average of the seven association values from the Latin American countries ($r = .36$).

3.4. Testing for systematic bias

Inspection of Table 1 reveals considerable variation in the accuracy of reporting among these studies. The number of studies that used categorical measures and presented cross-tabulation matrices ($k = 3$) was deemed insufficient to warrant a chi-square analysis of the number of participants who under-reported, over-reported, and accurately reported their objective behavior. However, a direct comparison between self-reports and objective measures to test for systematic under- or over-reporting was possible across the six studies that used a common metric. Analysis of the sample-size weighted self-report to objective measure ratios, averaged across the six studies, revealed slight

over-reporting; self-reports were, on average, 9% higher than objective measures.

3.5. Publication bias

An additional analysis was performed to assess the extent of publication bias. Rosenthal's (1979) classic fail-safe N was found to be 6065. Therefore, a very large number of studies with a mean effect size of zero would be necessary before the overall effect found in the present study would become statistically non-significant, which indicates that the present finding ($r = .46$) is very likely non-zero.

4. Discussion

Self-report measures of proenvironmental behavior are commonly used to reflect objective proenvironmental behavior, although to date, evidence for the validity of these measures has been mixed. This meta-analysis set out to investigate the degree of association between these two types of measures by combining the results of studies that contain both self-reports and objective measures of the same proenvironmental behavior(s). Across 6260 individuals and households, representing 19 association values from 15 studies, a large effect size ($r = .46$) was found, indicating that self-reported measures of proenvironmental behavior are, in general, highly associated with objective measures.

Based on Cohen's (1988) operational guidelines, correlation coefficients in the order of .50 are "large" in magnitude, making our observed mean effect size of $r = .46$ objectively large by definition. Furthermore, an effect size of this magnitude is unusually large compared to the typical effect sizes obtained in the broader social psychological literature. For example, Hemphill (2003) reanalyzed two extensive summaries of the psychological literature (Lipsey & Wilson, 1993; Meyer et al., 2001) and found that Cohen's benchmark for a large effect size corresponded to the 89th and 97th percentile, respectively, of studies included. In addition, Richard, Bond, and Stokes-Zoota (2003) analyzed 322 meta-analyses of a

¹¹ Without the inclusion of the Vadez et al. (2003) study on deforestation, this moderation no longer trends towards significance, $Q_{value(5)} = 5.50$, $p = .36$.

¹² The study conducted in Switzerland had the highest weighted correlation ($r = .81$), followed by Canada ($n = 3$; $r = .73$), Bolivia ($n = 1$; $r = .73$), Taiwan ($n = 2$; $r = .39$), Mexico ($n = 6$; $r = .33$), The Netherlands ($n = 1$; $r = .30$), and the United States ($n = 5$; $r = .29$).

Table 2Descriptive statistics for studies included in the analysis ($N = 15$).

Source	N	r value	Male (%)	Age (M)	Country
Bolderdijk et al. (2011)	138	.30	60.00	24.2	The Netherlands
Chao and Lam (2011)	172	.33	40.12	—	Taiwan
Corral-Verdugo (1997)	100	.31 – Re-use	0	44.00	Mexico
Corral-Verdugo (1997)	100	.14 – Recycling	0	44.00	Mexico
Corral-Verdugo (1995)	50	.08 – Re-use	0	—	Mexico
Corral-Verdugo (1995)	50	.25 – Recycling	0	—	Mexico
Corral-Verdugo (2003)	510	.38	40.35	36.6	Mexico
Corral-Verdugo and Figueredo (1999)	130	.63	30.77	35.2	Mexico
Cote (1984)	63 H ^a	.31	—	—	United States
Fuj et al. (1985)	2543H	.06	—	—	United States
Hamilton (1985)	471H	.18 ^b	—	—	United States
Hirst and Goeltz (1985)	196H	.63 ^c	—	—	United States
Kaiser et al. (2001)	40	.81	52.5	45.2	Switzerland
Lam and Cheng (2002)	200 ^d	.45	—	—	Taiwan
McGuire (1984)	73H	.16	—	—	United States
Vadez et al. (2003) ^e	24H	.73	100	—	Bolivia
Warriner et al. (1984)	969	.85 – Electricity usage	—	—	Canada
Warriner et al. (1984)	397	.58 – Gas usage	—	—	Canada
Warriner et al. (1984)	184	.70 – Oil usage	—	—	Canada

^a “H” indicates that households, rather than individuals, were the participants.^b Correlation calculated based on test of independence model: $\chi^2_{(6)} = 1.9, p = .01$.^c Calculated based on 2 x 2 odds ratio matrix (congruence between the number of retrofit measures reportedly installed, or not, by households and those observed, or not, by utility companies).^d Couples.^e Given that the nature of this study (i.e., deforestation) differs substantively from that of the other studies included, the analyses were run with and without this study, but the overall observed effect size remained largely unchanged. Moderator analyses conducted without this study are noted in the endnotes.

variety of social psychological phenomena and found that only 5% yielded an r value greater than .50, and also that none of the 18 social psychological topic areas included in the analysis yielded a mean correlation coefficient as high as .50. Thus, our observed overall effect size is both objectively large by definition and relatively unique in the size of its magnitude compared to the typical effect sizes obtained in the broader social psychological literature.

So, does this large effect size mean that self-reports are highly valid? Not necessarily. Two key reservations temper our enthusiasm to claim that our findings provide strong empirical evidence for the validity of self-report measures of proenvironmental behavior. First, and most importantly, 79% of the variance in the association between self-reported and objective behavior remains unexplained. This large amount of unexplained variance may not be

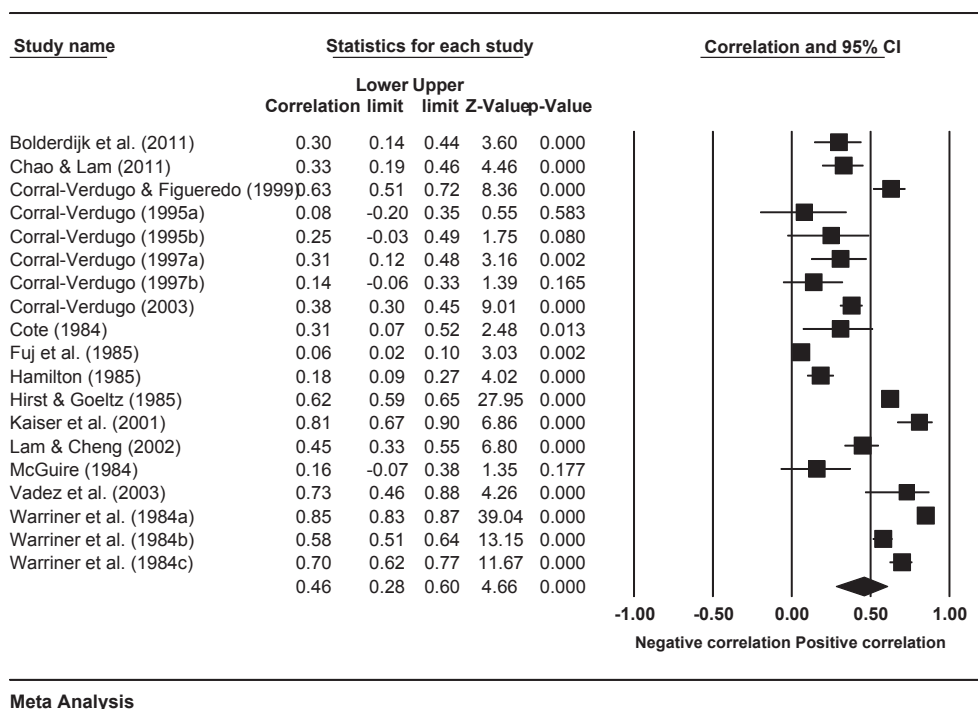


Fig. 1. Meta-analysis, using a random-effects model, of the standardized association between self-reported and objective measures of proenvironmental behavior for each article included in the analysis. The diamond represents the overall observed Pearson's r value.

as troubling in studies where the goal is to assess the *predictive* ability of certain variables or treatments; however, here the assumption is that self-report measures of proenvironmental behavior are a measure of the very behavior itself. In light of this, one might expect this percentage of explained variance to be considerably higher. Indeed, the guidelines for interpreting the magnitude of correlation coefficients are, to some extent, artificial and subjective, and so some have argued that the magnitude of an observed effect should be interpreted with respect to its context. Cohen himself states that, “The terms ‘small,’ ‘medium,’ and ‘large’ are relative, not only to each other, but to the area of behavioral science or even more particularly to the specific content and research method being employed in any given investigation ...” (1988, p. 25). The importance of a small effect size, for example, can become apparent in the context of a reduction in heart attack rates among doctors assigned to take aspirin versus those assigned to a placebo (Rosenthal & Rubin, 1979, 1982). Similarly, we would argue that an overall r of .46 becomes less impressive when we consider that self-reports are intended to be *de facto* measures of objective behavior. Second, the presence of highly heterogeneous findings reduces the estimated precision of the mean effect, and thus suggests the need for cautious interpretation of the mean effect size (Richard et al., 2003). In summary, given this particular context of self-report validity, we conclude that the observed effect size of $r = .46$ is conventionally large but functionally small.

4.1. Explaining observed heterogeneity

The validity of self-reports varied substantially across studies, from $r = .06$ to $r = .85$, suggesting that they are considerably more accurate in some situations than in others. It is not uncommon for social psychological effects to vary substantially from study to study (e.g., see Richard et al., 2003). In this analysis, two significant moderators influenced the strength of the observed association. First, studies with a higher percentage of males in the sample were found to be more valid. This is consistent with research demonstrating that females typically report more proenvironmental behavior than males, a tendency that may reduce report accuracy (e.g., Zelezny et al., 2000). Second, variation in the country location of the study also accounted for significant variation across studies, although it is difficult to interpret this finding and identify trends given the wide deviation in observed effect sizes across countries. Association values from non-Latin American countries ($r = .45$) were found to be slightly higher than those from the Latin American countries ($r = .36$). However, caution is required in speculating as to reasons for the difference in effect sizes between these two categories of countries, in part because the effect sizes from the Latin American countries include both the highest effect size obtained (i.e., $r = .73$ in Bolivia) as well as some of the lowest effect sizes obtained (e.g., $r = .08$ in Mexico). Furthermore, the difference may have occurred, for example, because of variations in the type of behavior under study (i.e., five of the seven association values from Latin America focused on waste and recycling behavior and the study with the highest self-report validity was unique in that it focused on deforestation behavior). Also, four of the five studies from Latin America were performed by the same researcher, and so it would seem unreasonable to draw conclusions about the entire region based on the available sample.

Importantly, the studies included in this analysis varied considerably from one another in terms of the types of measures employed and behaviors studied (e.g., water or energy use, reuse of materials, and food consumption), and the range of effect sizes was considerable across studies, indicating that self-report validity varied widely across the somewhat disparate studies examined. Although the type of behavior under study approached significance

as a moderating effect, additional future studies containing both types of measures are required for a fuller assessment of the influence of these various categories of estimates on effect size, thereby permitting us to evaluate the degree to which the observed heterogeneity is a result of what is being measured. Until the factors influencing self-report validity can be more fully elucidated, the present findings suggest the need for a cautious approach to interpreting the validity of results in the absence of a specific objective measure.

4.2. Systematic biases in self-reporting

Studies that used a common metric for their self-report and objective measure indicated a slight tendency toward over-reporting, which provides some support for the theoretical claim that self-reports may be exaggerated, possibly because they are subject to social desirability biases (see Stern & Oskamp, 1987; Tarrant & Cordell, 1997). An alternate explanation, of course, is that this slight tendency towards over-reporting may simply reflect individuals' lack of awareness about their actual proenvironmental behavior. In general, however, the wide variation shown in Table 1 suggests no overall systematic bias in reporting. This lack of overall systematic bias in reporting is consistent with recent findings that have shown low or no correlations between social desirability and self-reported proenvironmental behavior (e.g., Kaiser, Ranney, Hartig, & Bowler, 1999; Milfont, 2009; Schahn, 2002). Taken together, these results suggest that socially desirable responding does not exert a large influence on self-reported proenvironmental behavior. Of course, this conclusion must be tentative because of inconsistencies in study design, which made it difficult to draw meaningful, direct comparisons between self-reports and objective measures in many cases.

4.3. Limitations

Stringent inclusion criteria, set in place to ensure rigor of the analyses and meaningfulness of the results, meant that some relevant (and even seminal) studies were ineligible for inclusion in the analysis. These criteria also meant that limited data were available for the moderation analyses and the sample size was further limited in the case of some moderators where not every study contained the information (e.g., participant age) necessary for moderation analyses. However, it is not uncommon for meta-analyses with rigorous inclusion criteria to be limited to relatively small sample sizes (e.g., Hurst, Dittmar, Bond, & Kasser, 2013; Khemlani & Johnson-Laird, 2012; Stanton & Shadish, 1997; Van den Bussche, Van den Noortgate, & Reynvoet, 2009). There remains the possibility that, regardless of our attempts to thoroughly search for related articles, some manuscripts that would have been eligible for inclusion were inadvertently overlooked. We assume, however, that the vast majority of relevant studies would be included within the extensive PsycINFO database, as well as in our search of reference lists from relevant articles and review articles and a request for relevant published and unpublished articles on a popular subject listserve.

In addition, the moderating effects of two potentially important variables – social desirability bias and socioeconomic status (SES) – could not be explicitly examined because of insufficient information in the majority of sample studies. This is unfortunate given that recent findings (e.g., Milfont, 2009) have provided convincing evidence that contradicts the long-held assumption that self-reports of proenvironmental behavior are prone to influence by the social desirability bias (e.g., Stern & Oskamp, 1987; Sudman & Bradburn, 1974; Tarrant & Cordell, 1997); as such, it would have been ideal to explicitly evaluate the moderating effects of social desirability as

part of this meta-analysis. Also, the inability to assess SES is a major limitation given that previous research seems to suggest contradictory evidence about the influence of SES on self-report validity, with some studies on water conservation behavior finding that self-report accuracy is positively related to SES (Hamilton, 1985) and other studies on residential energy use finding that report accuracy is negatively related to family income (Warriner et al., 1984). It is unfortunate that this analysis could not assist in resolving some of these consistencies. This underscores the need for future research to report relevant SES data such that subsequent meta-analyses may examine its influence on self-report validity.

Another potential limitation is that the majority of the studies in this analysis involved *static* proenvironmental behavior, but two studies (i.e., Fuj et al., 1985; Hamilton, 1985) evaluated self-reported and objective *change* to electricity or water usage over the past year. This difference would not create any *statistical* problems because effect size values were all converted into standardized Fisher's *z* scores for analysis; however, there may be some potential *conceptual* implications of including studies that involve behavior change as well as those that examine static behavior. One can imagine that self-reporting change in proenvironmental behavior is different from self-reporting past and present proenvironmental behavior, and that this difference may translate into differences in validity. Unfortunately, we are unable to analyze this as a potential moderator given that there were only two studies that involved change.

Finally, the validity of the objective behavioral measures is another important potential limitation of the study. Intuitively it may seem as though certain measurements are individually 'objective' (e.g., instrument readings, archival data, behavioral observation, etc.) and yet, collectively, these measures can differ quite substantially from one another, likely yielding differing specific values. In the specific case of ratings from peers and trained observers, the line is blurrier than in the case of more classic objective measures, such as meter readings. One may be justly concerned that such ratings seem to have similar disadvantages as self-reported behavioral measures given that they are both subjective. Despite the potentially cloudy objectivity of several of the studies including peers and trained observers, these studies were included in this analysis because they are part of a relatively small number of studies that have attempted to assess self-report validity in relation to proenvironmental behavior and also because they are being used in the literature as quasi-objective measures to assess the accuracy of self-reports. In addition, the often private nature of proenvironmental behavior makes the peer-rating approach particularly well-suited to our research domain. Interestingly, research shows that self-reports and peer-reports, for example, diverge from one another (Chao & Lam, 2011; Lam & Cheng, 2002), but the validity of these measures is unknown and that inter-rater reliability is often absent. This highlights issues within the field concerning whether or not some of these measures that are being treated as objective are truly such. Given that objective measures themselves may be problematic as a reference variable given that they can be differently defined and operationally assessed, it seems desirable to strive for greater standardization in the objective measurement of behavior, similar to goals related to improving self-report measurement.

4.4. Future research

Future research should aim, in part, to identify the factors that help to explain the 79% of unaccounted-for variance between self-reports and their objective counterparts. Given that classical self-reports involve an assessment of subjective frequency, some researchers have argued that self-reports may reflect a self-

perception more than an estimate of actual behavior (Corral-Verdugo, 1997). Researchers may benefit by further exploring the differential ways in which environmental attitudes are related to both self-reported proenvironmental behavior and objective proenvironmental behavior.

Future research may also investigate the implications of the present results for theory development. Given that self-reports are not always highly valid, prediction models may inadvertently draw skewed conclusions about the usefulness of a theory in predicting objective behavior when the criterion variable is a self-report of proenvironmental behavior. As noted earlier, one meta-analysis found that the theory of planned behavior explained 10% more variance in self-reported behavior than in objective behavior (Armitage & Conner, 2001). Another study found that the theory of planned behavior predicted self-reported proenvironmental behavior considerably better than peer-reported proenvironmental behavior ($R^2 = .81$ and $R^2 = .10$, respectively; Chao & Lam, 2011). These results suggest that the validity of self-report measures should be considered in terms of their implications for theory development.

Carrying this further, the use of self-reports during the application phase may result in an overestimation of the influence of a behavioral intervention; for example, Porter, Leeming, and Dwyer (1995) found that the effect of recycling interventions tended to be stronger when researchers measured self-reported recycling, as opposed to the actual volume of recycled material. Therefore, problems can result from the development of theoretical models using self-reports as outcome measures when the intention is to later apply these models to real-world interventions. This mismatch between the type of behavior measured during the theory development phase and that targeted during the implementation phase presents a potential threat to intervention efficacy.

Finally, future research might well also attempt to identify the specific conditions (i.e., behavior and study design characteristics) under which self-reports have the highest and lowest validity, which would suggest to researchers when these measures may be most trusted. This meta-analysis identified some factors that moderate validity, and future research – with the benefit of a systematic program of research including prospective studies that contain *both* self-report and objective measures of various well-defined proenvironmental behaviors – could reveal other factors. For example, the type of behavior under study (which approached significance as a moderating effect in the present study) may emerge as a significant moderator with the benefit of additional future studies. The existence of such additional studies that include both measures would enable a subsequent meta-analysis with a larger sample size, which would allow for an exploration of interactive effects among the variables and a more detailed analysis of potential variables that moderate self-report validity.

4.5. Suggestions for researchers

Self-reports of retrospective behaviors can be strongly influenced by features of the research instrument (e.g., question format, wording, and context) such that minor changes in survey design can greatly influence results (cf. Schwarz, 1999). Based on results of the current study, as well as our experience and readings on self-report validity, we offer some recommendations for future research.

First, when researchers ask respondents to report the frequency with which they have engaged in a certain behavior over a specified reference period, they assume respondents search their memory for relevant episodes and merely sum them to arrive at a numeric response. Unfortunately, except for behaviors that are rare and

important, this model does not account for how participants respond to such frequency questions. Rather, respondents often base frequency estimates on fragmented recall and the application of inference rules (e.g., Sudman, Bradburn, & Schwarz, 1996). Everyday behaviors are poorly represented in memory, thus causing participants to depend on often flawed estimation strategies to arrive at frequency reports (e.g., Schwarz & Sudman, 1994). Moreover, features of the behavioral frequency question seem to greatly influence subsequent responses. For example, participants assume that the scale has been constructed such that its middle range reflects the 'average' or 'usual' behavioral frequency and, so, they should use the range of the response alternatives as a frame of reference in estimating their own behavioral frequency (cf. Schwarz, 1999). Participants may also assume that measures using a low-frequency scale (e.g., "number of times per week") refer to major occurrences and those with high-frequency reports (e.g., "number of times per day") refer to minor occurrences (Schwarz, Strack, Müller, & Chassein, 1988). The length of the reference period (e.g., referring to "last week" versus "last year") can also have a similar effect on question interpretation (Winkielman, Knäuper, & Schwarz, 1998).

These systematic influences of response alternatives can be minimized (Schwarz, 1999). The accuracy of responses to behavioral frequency questions is greater when these questions are asked in an open-response format (e.g., "How many minutes a day do you spend in the shower? ____"), with relevant units of measurement specified (i.e., minutes per day) because this format includes less researcher-induced measurement error to distort the responses (e.g., there is no rating scale that requires the researcher to develop verbal phrases to label the points; Krosnick, Judd, & Wittenbrink, 2005). Where possible, researchers should avoid using vague quantifiers, such as 'sometimes' and 'often,' to assess objective frequencies of behavior (cf. Moxey & Sanford, 1992) because these terms can carry different meanings in different content domains and also to different participants.

Krosnick et al. (2005) support the use of 7-point scales, and suggest that scale points must be perceived by participants to be ordinal, progressing meaningfully from one end of the scale to the other, and that the meaning of adjacent points should not overlap with one another. Respondents should have a relatively precise and stable understanding and interpretations of the meaning of each scale point. All scale points should be labeled with words as opposed to just some scale points being labeled (e.g., Krosnick & Berent, 1993), and labels should be selected such that they divide the continuum into equal units in order to maximize validity (e.g., Klockars & Yamagishi, 1988). Last, self-report data is best collected anonymously and via self-administered questionnaires to reduce the potential for social pressure and the possibility of social desirability biases.

4.6. Conclusion

This meta-analysis revealed a strong association between self-reported and objective proenvironmental behavior. Nevertheless, the considerable amount of unexplained variance between self-reports and objective measures suggests that these measures are not as isomorphic as some might think, or expect, given the particular context of self-report validity, which leads us to conclude that the observed effect size is conventionally large but functionally small. This is not to argue that self-reports should not be used. Far from it, in fact, they are convenient and cost-effective indicators of behavior and they have high levels of validity in some cases. Furthermore, the constraints of some research studies (e.g., budgetary and time-wise) dictate the preferred use of self-reports over objective measures. Rather, this research highlights the fact that self-report validity can

vary widely across studies and so calls for caution in interpretation in studies based on self-report measures.

Ideally we would have been able to include many more studies in this analysis, would not have observed significant heterogeneity among the included studies, and all of the measures employed in the studies would have been undeniably 'objective.' But these are all issues that exist in the field and, as such, aside from the meta-analytic findings themselves, the shortcomings of the manuscript are, in and of themselves, also informative because they highlight areas that would benefit from further research attention and discussion within the field. Given these limitations, however, there is a need for cautious interpretation of the present findings, especially because of the observed variations in effect sizes as well as study design.

Because the use of self-report measures is seemingly unavoidable, research related to self-report validity is important. Indeed, if our ultimate goal is to help effect *objective behavior change*, then we have much to gain by focusing on establishing and improving self-report validity, as well as including more objective measures, for the purposes of assessing proenvironmental behavior.

Acknowledgments

The first author wishes to gratefully acknowledge funding support received from the Social Sciences and Humanities Research Council (Joseph-Armand Bombardier Canada Graduate Scholarships – Doctoral award (Award No. 767-2010-1141)), as well as the Pacific Institute for Climate Solutions. In addition, the authors would like to acknowledge the research assistance of Emma Fraser and Jas Johal.

References¹³

- Affuso, O., Stevens, J., Catellier, D., McMurray, R. G., Ward, D. S., Lytle, L., et al. (2011). Validity of self-reported leisure-time sedentary behavior in adolescents. *Journal of Negative Results in Biomedicine*, 10, 2. <http://dx.doi.org/10.1186/1477-5751-10-2>.
- Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl, & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 11–39). New York: Springer-Verlag.
- Ajzen, I. (2005). *Attitudes, personality and behavior*. Maidenhead: McGraw-Hill International (UK) Ltd.
- Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behavior: A meta-analytic review. *British Journal of Social Psychology*, 40, 471–499. <http://dx.doi.org/10.1348/014466601164939>.
- Barr, S. (2007). Factors influencing environmental attitudes and behaviours: A U.K. case study of household waste management. *Environment and Behavior*, 39, 435–473. <http://dx.doi.org/10.1177/0013916505283421>.
- Boldero, J. (1995). The prediction of household recycling of newspapers: The role of attitudes, intentions, and situational factors. *Journal of Applied Social Psychology*, 25, 440–462. <http://dx.doi.org/10.1111/j.1559-1816.1995.tb01598.x>.
- *Bolderdijk, J. W., Knockaert, J., Steg, E. M., & Verhoef, E. T. (2011). Effects of Pay-As-You-Drive vehicle insurance on young drivers' speed choice: Results of a Dutch field experiment. *Accident Analysis & Prevention*, 43, 1181–1186. <http://dx.doi.org/10.1016/j.aap.2010.12.032>.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis version 2*. Englewood, NJ: Boistat.
- Brener, N. D., Billy, J. O. G., & Grady, W. R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health*, 33, 436–457. [http://dx.doi.org/10.1016/S1054-139X\(03\)00052-1](http://dx.doi.org/10.1016/S1054-139X(03)00052-1).
- *Chao, Y.-L., & Lam, S.-P. (2011). Measuring responsible environmental behavior: Self-reported and other-reported measures and their differences in testing a behavioral model. *Environment & Behavior*, 43, 53–71. <http://dx.doi.org/10.1177/0013916509350849>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. <http://dx.doi.org/10.1037/0033-2909.112.1.155>.

¹³ Note: References denoted with an asterisk (*) are those studies included in our meta-analysis.

- *Corral-Verdugo, V. (1997). Dual "realities" of conservation behavior: Self-reports vs. observations of reuse and recycling behavior. *Journal of Environmental Psychology*, 17, 135–145. <http://dx.doi.org/10.1006/jenvp.1997.0048>.
- *Corral-Verdugo, V., Bernache, G., Encinas, L., & Garibaldi, L. C. (1995). A comparison of two measures of reuse and recycling behavior: Self-report and material culture. *Journal of Environmental Systems*, 23, 313–327.
- *Corral-Verdugo, V., Bechtel, R. B., & Fraijo-Sing, B. (2003). Environmental beliefs and water conservation: An empirical study. *Journal of Environmental Psychology*, 23, 247–257. [http://dx.doi.org/10.1016/S0272-4944\(02\)00086-5](http://dx.doi.org/10.1016/S0272-4944(02)00086-5).
- *Corral-Verdugo, V., & Figueredo, A. J. (1999). Convergent and divergent validity of three measures of conservation behavior: The multitrait-multimethod approach. *Environment & Behavior*, 31, 805–820. <http://dx.doi.org/10.1177/00139169921972353>.
- *Cote, J. A. (1984). Use of household refuse analysis to measure usual and period-specific food consumption. *American Behavioral Scientist*, 28, 129–138. <http://dx.doi.org/10.1177/000276484028001010>.
- De Young, R. (1990). Promoting conservation behavior in shared spaces: The role of energy monitors. *Journal of Environmental Systems*, 19, 265–273.
- Denny, S. J., Milfont, T. L., Utter, J., Robinson, E. M., Ameratunga, S. N., Merry, S. N., et al. (2008). Hand-held internet tablets for school-based data collection. *BMC Research Notes*, 1, 52. <http://dx.doi.org/10.1186/1756-0500-1-52>.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Holt, Rinehart and Winston.
- Elgar, F. J., & Stewart, J. M. (2008). Validity of self-report screening for overweight and obesity: Evidence from the Canadian community health survey. *Canadian Journal of Public Health*, 99, 423–427. Retrieved from <http://search.proquest.com.ezproxy.library.uvic.ca/docview/232005343?accountid=14846>.
- Evans, G. W. (1999). Measurement of the physical environment as stressor. In S. L. Friedman, & T. D. Wachs (Eds.), *Assessment of the environment across the lifespan* (pp. 249–277). Washington, DC: American Psychological Association.
- *Fuj, E. T., Hennessy, M., & Mak, J. (1985). An evaluation of the validity and reliability of survey response data on household electricity conservation. *Evaluation Review*, 9, 93–104. <http://dx.doi.org/10.1177/0193841X8500900106>.
- Gatersleben, B., Steg, L., & Vlek, C. (2002). Measurement and determinants of environmentally significant consumer behavior. *Environment and Behavior*, 34, 335–362. <http://dx.doi.org/10.1177/0013916502034003004>.
- Geller, E. S. (1981). Evaluating energy conservation programs: Is verbal report enough? *Journal of Consumer Research*, 8, 331–335. Retrieved from <http://www.jstor.org/stable/2488892>.
- Gifford, R. (2006). A general model of social dilemmas. *International Journal of Ecological Economics and Statistics*, 5, 23–40.
- *Hamilton, L. C. (1985). Self-reported and actual savings in a water conservation campaign. *Environment & Behavior*, 17, 315–326. <http://dx.doi.org/10.1177/0013916585173003>.
- *Hirst, E., & Goeltz, R. (1985). Accuracy of self-reports: Energy conservation surveys. *The Social Science Journal*, 22, 19–30.
- Heath, Y., & Gifford, R. (2002). Extending the theory of planned behavior: Predicting the use of public transportation. *Journal of Applied Social Psychology*, 32, 2154–2185. <http://dx.doi.org/10.1111/j.1559-1816.2002.tb02068.x>.
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, 58, 78–80. <http://dx.doi.org/10.1037/0003-066X.58.1.78>.
- Hurst, M., Dittmar, H., Bond, R., & Kasser, T. (2013). The relationship between materialistic values and environmental attitudes and behaviours: A meta-analysis. *Journal of Environmental Psychology*, 36, 257–269. <http://dx.doi.org/10.1016/j.jenvp.2013.09.003>.
- Kaiser, F. G., Doka, G., Hofstetter, P., & Ranney, M. A. (2003). Ecological behavior and its environmental consequences: A life cycle assessment of a self-report measure. *Journal of Environmental Psychology*, 23, 11–20. [http://dx.doi.org/10.1016/S0272-4944\(02\)00075-0](http://dx.doi.org/10.1016/S0272-4944(02)00075-0).
- Kaiser, F. G., Ranney, M., Hartig, T., & Bowler, P. A. (1999). Ecological behavior, environmental attitude, and feelings of responsibility for the environment. *European Psychologist*, 4, 59–74.
- *Kaiser, F. G., Frick, J., & Stoll-Kleemann, S. (2001). Zur Angemessenheit selbstberichteten Verhaltens: Eine Validitätsuntersuchung der Skala Allgemeinen Ökologischen Verhaltens [Accuracy of Self-Reports: Validating the General Ecological Behavior Scale]. *Diagnostica*, 47, 88–95.
- Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, 138, 427–457. <http://dx.doi.org/10.1037/a0026841>.
- Klockars, A. J., & Yamagishi, M. (1988). The influence of labels and positions in rating scales. *Journal of Educational Measurement*, 25, 85–96. <http://dx.doi.org/10.1111/j.1745-3984.1988.tb00294.x>.
- Krosnick, J. A., & Berent, M. K. (1993). Comparisons of party identification and policy preferences: The impact of survey question format. *American Journal of Political Science*, 37, 941–964. <http://dx.doi.org/10.2307/2111580>.
- Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). The measurement of attitudes. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The Handbook of attitudes* (pp. 21–78). Mahwah, NJ: Lawrence Erlbaum Associates.
- *Lam, S.-P., & Cheng, S.-I. (2002). Cross-informant agreement in reports of environmental behavior and the effect of cross-questioning on report accuracy. *Environment and Behavior*, 34, 508–520. <http://dx.doi.org/10.1177/00116502034004006>.
- Lee, Y.-J. (1993). Recycling behavior and waste management planning. *Journal of Building and Planning National Taiwan University*, 7, 65–77.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181–1209. <http://dx.doi.org/10.1037/0003-066X.48.12.1181>.
- Luyben, P. D. (1982). Prompting thermostat setting behavior. *Environment and Behavior*, 14, 113–128. <http://dx.doi.org/10.1177/0013916582141007>.
- McCormack, G. R., Cerin, E., Leslie, E., du Toit, L., & Owen, N. (2008). Objective versus perceived walking distances to destinations: Correspondence and predictive validity. *Environment and Behavior*, 40, 401–425. <http://dx.doi.org/10.1177/0013916507300560>.
- *McGuire, R. H. (1984). Recycling, great expectations and garbage outcomes. *American Behavioral Scientist*, 28, 71–91. <http://dx.doi.org/10.1177/000276428402800107>.
- Margai, F. L. (1997). Analyzing changes in waste reduction behavior in a low-income urban community following a public outreach program. *Environment and Behavior*, 29, 769–792. <http://dx.doi.org/10.1177/0013916597296003>.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128–165. <http://dx.doi.org/10.1037/0003-066X.56.2.128>.
- Milfont, T. L. (2009). The effects of social desirability on self-reported environmental attitudes and ecological behaviour. *The Environmentalist*, 29, 263–269. <http://dx.doi.org/10.1007/s10669-008-9192-2>.
- Moxey, L. M., & Sanford, A. J. (1992). Context effects and the communicative functions of quantifiers: Implications for their use in attitude research. In N. Schwarz, & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 279–296). New York: Springer Verlag.
- Nency-Thiel, M., Beal, V., Ludwischowska, G., & Romaniuk, J. (2013). Investigating the accuracy of self-reports of brand usage behavior. *Journal of Business Research*, 66, 224–232. <http://dx.doi.org/10.1016/j.jbusres.2012.07.016>.
- Olson, M. E. (1981). Consumers attitudes toward energy conservation. *Journal of Social Issues*, 37, 108–131.
- Peterson, R. A., & Kerin, R. (1981). The quality of self-report data: Review and synthesis. In B. Enis, & K. Roering (Eds.), *Annual review of marketing* (pp. 5–20). Chicago: American Marketing Association.
- Porter, B. E., Leeming, F. C., & Dwyer, W. O. (1995). Solid waste recovery: A review of behavioral programs to increase recycling. *Environment & Behavior*, 27, 122–152. <http://dx.doi.org/10.1177/0013916595272002>.
- Rathje, W. L. (1989). The three faces of garbage – Measurements, perceptions, behaviors. *Journal of Management and Technology*, 17, 61–65.
- Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7, 331–363. <http://dx.doi.org/10.1037/1089-2680.7.4.331>.
- Rolnick, S. J., Parker, E. D., Nordin, J. D., Hedblom, B. D., Wei, F., Kerby, T., et al. (2013). Self-report compared to electronic medical record across eight adult vaccines: Do results vary by demographic factors? *Vaccine*, 31, 3928–3935. <http://dx.doi.org/10.1016/j.vaccine.2013.06.041>.
- Rosenthal, R. (1979). The 'file drawer problem' and tolerance for null results. *Psychological Bulletin*, 86, 638–641. doi: 0.1037/0033-2909.86.3.638.
- Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Psychology*, 9, 395–396. <http://dx.doi.org/10.1111/j.1559-1816.1979.tb02713.x>.
- Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166–169. <http://dx.doi.org/10.1037/0022-0663.74.2.166>.
- Schahn, J. (2002). The role of behavioral costs and social desirability as predictors of environmental attitudes and conservation behavior: An analysis on aggregate and on individual data level. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 23, 45–54.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54, 93–105. <http://dx.doi.org/10.1037/0003-066X.54.2.93>.
- Schwarz, N., & Sudman, S. (1994). *Autobiographical memory and the validity of retrospective reports*. New York: Springer Verlag.
- Schwarz, N., Strack, F., Müller, G., & Chassein, B. (1988). The range of response alternatives may determine the meaning of the question: Further evidence on informative functions of response alternatives. *Social Cognition*, 6, 107–117. <http://dx.doi.org/10.1521/soco.1988.6.2.107>.
- Stanton, M. D., & Shadish, W. R. (1997). Outcome, attrition, and family-couples treatment for drug abuse: A meta-analysis and review of the controlled, comparative studies. *Psychological Bulletin*, 122, 170–191. <http://dx.doi.org/10.1037/0033-2909.122.2.170>.
- Steg, L., & Vlek, C. (2009). Encouraging pro-environmental behaviour: an integrative review and research Agenda. *Journal of Environmental Psychology*, 29, 309–317.
- Stern, P. C., Dietz, T., Kalof, L., & Guagnano, G. A. (1995). Values, beliefs, and pro-environmental action: Attitude formation toward emergent attitude objects. *Journal of Applied Social Psychology*, 25, 1611–1636. <http://dx.doi.org/10.1111/j.1559-1816.1995.tb02636.x>.
- Stern, P. C., & Oskamp, S. (1987). Managing scarce environmental resources. In D. Stokols, & I. Altman (Eds.), *Handbook of environmental psychology* (pp. 1043–1088). New York: Wiley.
- Stutts, J. L., Ghate, S. R., Gill, J. L., Stutts, C. R., Barnes, C. N., Lajoie, A. S., et al. (2006). Validity of self-reported smoking status among participants in a lung cancer screening trial. *Cancer Epidemiology, Biomarkers & Prevention*, 15, 1825–1828. <http://dx.doi.org/10.1158/1055-9965.EPI-06-0393>.
- Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys*. Chicago: Aldine Publishing Company.

- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.
- Tarrant, M. A., & Cordell, H. K. (1997). The effect of respondent characteristics on general environmental attitude-behavior correspondence. *Environment & Behavior*, 29, 618–637. <http://dx.doi.org/10.1177/0013916597295002>.
- *Vadez, V., Reyes-Garcia, V., Godoy, R., Williams, L., Apaza, L., Byron, E., et al. (2003). Validity of self-reports to measure deforestation: Evidence from the Bolivian Lowlands. *Field Methods*, 15, 289–304. <http://dx.doi.org/10.1177/1525822X03254847>.
- Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin*, 135, 452–477. <http://dx.doi.org/10.1037/a0015329>.
- Van Liere, K. D., & Dunlap, R. E. (1980). The social bases of environmental concern: A review of hypotheses, explanations, and empirical evidence. *Public Opinion Quarterly*, 44, 181–197. Retrieved from <http://www.jstor.org/stable/2748427>.
- *Warriner, G. K., McDougall, G. H. C., & Claxton, J. D. (1984). Any data or none at all? Living with inaccuracies in self-reports of residential energy consumption. *Environment & Behavior*, 16, 503–526. <http://dx.doi.org/10.1177/0013916584164005>.
- Weinhardt, L. S., Forsyth, A. D., Carey, M. P., Jaworski, B. C., & Durant, L. E. (1998). Reliability and validity of self-report measures of HIV-related sexual behaviour: Progress since 1990 and recommendations for research and practice. *Archives of Sexual Behavior*, 27, 155–180. <http://dx.doi.org/10.1023/A:1018682530519>.
- Winkielman, P., Knäuper, B., & Schwarz, N. (1998). Looking back at anger: Reference periods change the interpretation of (emotion) frequency questions. *Journal of Personality and Social Psychology*, 75, 719–728. <http://dx.doi.org/10.1037/0022-3514.75.3.719>.
- Zelezny, L. C., Chua, P.-P., & Aldrich, C. (2000). Elaborating on gender differences in environmentalism. *Journal of Social Issues*, 56, 443–457. <http://dx.doi.org/10.1111/0022-4537.00177>.