

Special issue: Methodological Innovations: Original Research Report

Extracting Response Style Bias From Measures of Positive and Negative Affect in Aging Research

Stefan Schneider

Dornsife Center for Self-Report Science, Center for Economic and Social Research, University of Southern California, Los Angeles.

Correspondence should be addressed to Stefan Schneider, PhD, Dornsife Center for Self-Report Science, Center for Economic and Social Research, University of Southern California, 635 Downey Way, Los Angeles, CA 90089-3332. E-mail: schneids@usc.edu

Received February 8, 2016; Accepted July 31, 2016

Decision Editor: Shevaun Neupert, PhD

Abstract

Objectives: This study investigated the role of response style biases in the assessment of positive and negative affect in aging research; it addressed whether response styles (a) are associated with age-related changes in cognitive abilities, (b) lead to distorted conclusions about age differences in affect, and (c) reduce the convergent and predictive validity of affect measures in relation to health outcomes.

Method: A multidimensional item response theory model was used to extract response styles from affect ratings provided by respondents to the psychosocial questionnaire ($n = 6,295$; aged 50–100 years) in the Health and Retirement Study (HRS).

Results: The likelihood of extreme response styles (disproportionate use of “not at all” and “very much” response categories) increased significantly with age, and this effect was mediated by age-related decreases in HRS cognitive test scores. Removing response styles from affect measures did not alter age patterns in positive and negative affect; however, it consistently enhanced the convergent validity (relationships with concurrent depression and mental health problems) and predictive validity (prospective relationships with hospital visits, physical illness onset) of the affect measures.

Discussion: The results support the importance of detecting and controlling response styles when studying self-reported affect in aging research.

Keywords: Affect measurement—Extreme response bias—Multidimensional IRT

The assessment of affective and emotional experiences plays a vital role in aging research, both for capturing life-span changes in psychological well-being (Carstensen, Pasupathi, Mayr, & Nesselrode, 2000; Carstensen et al., 2011; Stone, Schwartz, Broderick, & Deaton, 2010) and for the prediction of physical mobility, morbidity, and mortality (Charles & Carstensen, 2010; Lyra, Törmäkangas, Read, Rantanen, & Berg, 2006; Ong, 2010; Steptoe & Wardle, 2011). Although several methods of affect measurement exist (e.g., autonomic nervous system responses, neuroimaging, facial expressions; Mauss & Robinson, 2009), the predominant approach is via self-report (Ekkekakis, 2013). Self-reports, however,

are potentially susceptible to various response biases and differences in how individuals use the response scale (Schwarz & Knäuper, 2012). One type of response bias, commonly referred to as response style, describes a person's tendency to use the rating scale in a certain systematic or “stylistic” way that is unrelated to the content of the items (Paulhus, 1991). Although research on response styles has a long history (Cronbach, 1950), these effects have found relatively little attention in the socioemotional aging literature. Thus, the purpose of this methodological study was to investigate evidence for the importance of considering response styles in the measurement of affective experiences in aging research.

Prominent types of response styles discussed in the literature include respondents' preference for using the most positive scale end point ("acquiescence"), the most negative scale end point ("disacquiescence"), or both extreme scale end points ("extreme responding"; Van Vaerenbergh & Thomas, 2013). Whether these response styles differ by age has been examined in several studies, predominantly in the areas of personality and marketing research (Van Vaerenbergh & Thomas, 2013). The available evidence suggests that acquiescent responding increases with age (Billiet & McClendon, 2000; Morales-Vives, Vigil-Colet, Lorenzo-Seva, & Ruiz-Pamies, 2014; Weijters, Geuens, & Schillewaert, 2010), and one study found a trend for greater disacquiescence with higher age (Weijters et al., 2010). Age effects for extreme response styles have been more mixed, with some studies documenting age-related increases (Greenleaf, 1992; Kieruj & Moors, 2013; Meisenberg & Williams, 2008; Weijters et al., 2010), and others finding no (Johnson, Kulesa, Cho, & Shavitt, 2005; Moors, 2008) or negative relationships (Austin, Deary, & Egan, 2006) with age. In a study of 12,000 adults, De Jong, Steenkamp, Fox, and Baumgartner (2008) found a curvilinear age-effect that might explain the results for extreme response style, in that extreme responding decreased from younger to middle age and increased from middle to old age.

Although these studies suggest that response style tendencies may generally increase in later parts of the adult life span, they only provide a fragmented picture of the importance of response style effects in aging research on affect and emotion. First, to date studies have been limited to asking whether age differences in response styles exist, but we know little about why they exist. Evidence of mechanisms underlying age differences in response styles could facilitate strategies to avoid or reduce such effects. Second, it remains an open question whether response styles actually translate into meaningful biases in affect measurement. If response styles change with age, this may undermine the accuracy of findings on age-related changes in emotional experiences and may diminish the concurrent or predictive validity of affect self-reports in aging research. These topics were addressed in the present study.

Cognitive Skills and Response Style Behaviors

A potentially important factor that may underlie age-related changes in response style behaviors is participants' cognitive skills. Theories on the psychology of survey methods agree that answering survey questions requires considerable cognitive effort: Respondents are expected to interpret the meaning of each question, retrieve relevant information from memory, integrate it into a summary judgment, and map their judgment onto the provided response alternatives (Schwarz & Knäuper, 2012; Tourangeau, Rips, & Rasinski, 2000). Krosnick (1991) posits that when the cognitive demands of accurately completing questionnaires challenge or exceed a respondent's cognitive abilities,

response style behaviors are more likely to occur. That is, to the extent that the cognitive costs of working through a series of questions becomes frustrating or burdensome, respondents may shift their response strategy from carefully considering the details of each item to more expedient and potentially "stylistic" selection of response choices. Robust evidence suggests that although there are minimal age-associated changes in some areas of cognitive functioning (e.g., crystallized verbal abilities), a number of cognitive abilities decrease in older age (McArdle, Fisher, & Kadlec, 2007; Wilson et al., 2002). Cognitive functions that show age-related declines include reductions in perceptual and mental processing speed, working memory, and episodic memory (Salthouse & Ferrer-Caja, 2003; Wilson et al., 2002), all of which are likely relevant for the task of completing a questionnaire. Accordingly, the present study examined the hypothesis that lower cognitive processing skills are associated with greater response style behavior and that this contributes to (i.e., mediates) age-related differences in response styles.

Implications of Response Styles for Research on Affect and Aging

Response styles have been regarded as a "nuisance" factor underlying the observed ratings that interfere with the measurement of the intended construct (Podsakoff, MacKenzie, & Podsakoff, 2012). As such, response styles can be viewed as introducing (random or systematic) error that decreases the validity of affect measures. By the same logic, removing response style effects should increase the validity of the scale scores. This conjecture was tested in the present study in several ways.

My first question was whether response styles distort age-related changes in positive and negative affect levels. Researchers have often been puzzled by the "well-being paradox of aging" (Swift et al., 2014), with repeated findings showing that negative affect decreases with age and that positive affect is maintained at high levels at least until after adults reach 70 or 80 years of age (Carstensen et al., 2000, 2011; Stone et al., 2010). These findings have been attributed to better emotion regulation and greater maturity in older age, but they have also raised concerns about measurement artifacts, including age-related biases from selective memory (Charles et al., in press), tendencies for socially desirable responding (Soubelet & Salthouse, 2011), and interviewer demand characteristics (Luong, Charles, Rook, Reynolds, & Gatz, 2015). If response styles differ by participant age, this may undermine comparisons of affect across age groups and may bias conclusions about developmental trajectories in affective well-being.

My second question was whether adjusting affect measures for response styles increases their convergent and predictive validity in aging research. Growing evidence supports the notion that both positive and negative emotions play important roles in the etiology of illness, medical service utilization,

and other health outcomes in older adulthood (Charles & Carstensen, 2010; Ong, 2010). If affect measures are biased by response styles, this may undermine their ability to predict concurrent or future health outcomes and it may put the predictive effects of self-report variables at a disadvantage when compared with the effects of other variables that are not based on self-report (e.g., objective biological indicators).

In summary, the current study builds on the literature by examining whether (a) age-related increases in response styles are evident in measures of positive and negative affect, (b) differences in cognitive abilities partially explain age differences in response styles, and (c) correcting self-reports for response styles accentuates (or attenuates) age differences in affective well-being, and improves the convergent and predictive validity of positive and negative affect measures. The predictions were examined using data from the Health and Retirement Study (HRS), a nationally representative study of adults aged 50 and older.

Method

Data and Sample

The HRS is a biennial panel survey of older Americans and their spouses that started in 1992 (<http://hrsonline.isr.umich.edu/>). The 2008 survey introduced measures of positive and negative affect (piloted in 2006), administered to a random 50% of the HRS sample who were selected to undergo an enhanced face-to-face interview. At the end of the interview, respondents were given a paper-and-pencil psychosocial questionnaire package to be returned in the mail, with a response rate of 89% (Smith et al., 2013). The present analyses included all respondents from the 2008 wave who were (a) 50 years of age or older, (b) eligible for the psychosocial questionnaires, and (c) completed the questions by themselves (excluded were about 2% of the questionnaires completed by proxy respondents), with a resulting sample size of 6,295 respondents. Descriptive sample characteristics are shown in Table 1.

Measures

Positive and negative affect

Positive affect was measured with 13 items and negative affect with 12 items from the Positive and Negative Affect Schedule—Expanded Form (Watson & Clark, 1994), supplemented by items from Carstensen and colleagues (2000). Respondents were asked “During the last 30 days, to what degree did you feel (afraid, upset, determined, enthusiastic, etc.)?”, with response options Very much, Quite a bit, Moderately, A little, Not at all. The affect questions were administered about halfway through the psychosocial questionnaire package (with 26 other questionnaires preceding), with positive and negative affect items presented in mixed order. Positive and negative affect sum scores were correlated at $r = -.38$, with Cronbach’s alphas of .89 for negative and .92 for positive affect.

Table 1. Descriptive Sample Characteristics

	Frequency (%)	N
Age (mean, <i>SD</i> /range)	69.82 (9.7/50–100)	6,295
Years of education (mean, <i>SD</i> /range)	12.59 (3.1/0–17)	6,288
Female	3,766 (59.8)	6,295
White race	5,243 (83.3)	6,294
Hispanic	538 (8.6)	6,294
Married/living together	4,092 (65.0)	6,294
Cognitive ability score (mean, <i>SD</i> /range)	0.66 (0.6/0–1)	6,257
CES-D score (mean, <i>SD</i> /range)	1.36 (1.9/0–8)	6,256
CIDI-SF major depression	264 (4.2)	6,257
Diagnosis of mental health problems	1,083 (17.4)	6,213
New onset of physical illness over next 4 years ^a	1,957 (36.7)	5,339
Hospital stay over next 4 years ^a	2,480 (46.0)	5,389

Notes: CES-D = Center for Epidemiologic Studies Depression Scale; CIDI-SF = Composite International Diagnostic Interview—Short Form.

^aData on illness onset and hospital stays are aggregated from 2010 and 2012 interviews. All other variables are from the 2008 interview.

Cognitive tests

Cognitive tests were administered in the core HRS interview. Using the imputed cognitive data set created by Fisher, Hassan, Faul, Rodgers, and Weir (2015), an average score was created from four tasks: (a) immediate free recall of a list of 10 nouns, (b) delayed recall of the same nouns 5 min later, (c) counting backward from 20, and (d) “serial 7s,” that is, counting backward from 100 by 7s. Each test score was transformed into a proportion correct score before averaging into the composite measure (higher scores represent better cognitive abilities); Cronbach’s alpha was .66. Additional cognitive tests in the HRS (person and object naming, providing today’s date) were not included here because they were administered only to a subset of older participants (Fisher et al., 2015).

Convergent validity measures

Three mental health measures assessed in the 2008 core interview were used to examine the convergent validity of the affect scores: (a) the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977), (b) the Composite International Diagnostic Interview—Short Form (CIDI-SF), and (c) diagnosis of mental health problems.

The CES-D is a widely used screening measure of depressive symptoms in the general population (Nezu, Nezu, McClure, & Zwick, 2002). It measures a continuum of psychological distress in the past week, with higher scores indicating higher symptom severity. The HRS uses an abbreviated eight-item version with dichotomous (yes/no) response format, which has demonstrated adequate construct validity (Steffick, 2000). Cronbach’s alpha was .80 in this sample.

The CIDI-SF is a descendent of the Diagnostic Interview Schedule designed to evaluate the presence of major depression over the year preceding the interview. In accordance

with DSM criteria, participants had to report depressed feelings or loss of interest for at least two consecutive weeks plus a total of five or more symptoms to meet the cutoff for major depression (Steffick, 2000).

To assess whether participants had a diagnosis of mental health problems, they were asked to report whether a doctor had ever told them that they had “emotional, nervous, or psychiatric problems.”

Predictive validity measures

Two measures of prospective physical health outcomes examined the predictive validity of affect: (a) new onset of a chronic physical illness and (b) hospital stays. Both health outcomes were assessed using data from the 2010 and 2012 waves following the 2008 affect assessment.

For the onset of physical illness, respondents were asked whether a doctor had ever told them that they had high blood pressure, diabetes, cancer, lung disease, heart disease, stroke, and arthritis. The variable was coded “yes” if any illness was new (i.e., not reported up to 2008 and reported for the first time in 2010 or 2012) and “no” otherwise.

For hospital stays, respondents were asked whether they had any overnight hospital stay since the last interview. The variable was coded dichotomously as “yes” if a hospital stay was indicated in the 2010 or 2012 wave and “no” otherwise.

Data Analysis

Measurement of response styles

A number of strategies for assessing response styles have been proposed in the literature (see Van Vaerenbergh & Thomas, 2013). Traditional methods use a simple sum-score index counting how often a participant has selected a particular (e.g., extreme) response category. A major limitation of this index is that it confounds the measurement of response styles and substantive content. For example, a person providing many high scores on a set of negative affect items could be an acquiescent responder, but could also be actually very high in negative affect. A simple sum-score index cannot disentangle these effects, which limits its usefulness: It commonly requires designated items for response style measurement and it does not provide a direct way of correcting the substantive measure for response style bias (Bolt, Lu, & Kim, 2014).

Recently proposed analytic methods address these problems by incorporating the measurement of both substantive and response style factors simultaneously in a single latent variable model. The approach applied in this study uses a multidimensional extension of the nominal response model (using item response theory [IRT]). A key feature of this IRT model is that the response choices are treated as unordered (nominal) categories; this makes it possible to differentiate substantive (i.e., positive or negative affect) and “stylistic” response components. Response style factors capture a person’s tendency to choose a particular response

option (e.g., the probability of choosing extreme response categories) regardless of the person’s level on the substantive affect factor, and the person’s affect level is estimated controlling (i.e., corrected) for the extracted response styles in the same model. The usefulness of this approach for detecting and correcting response styles has previously been documented (Bolt & Johnson, 2009; Falk & Cai, *in press*; Morren, Gelissen, & Vermunt, 2011). The Supplementary Material provides details on the model and its implementation using Mplus software in the present study.

Following recommended procedures (Morren et al., 2011), a stepwise model selection approach was used to determine the types of response styles (if any) present in the data. In a first step, an unrestricted (i.e., exploratory) response style factor was added to the substantive (affect) factor in which the response probabilities could take any shape. The dominant pattern detected in this step informed subsequent models testing specific prespecified response styles (e.g., extreme responses). The fit of different models was compared using the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for model selection.

Positive and negative affect items were examined separately. Following prior research (Billiet & McClendon, 2000; Kieruj & Moors, 2013; Moors, 2008; Morren et al., 2011), response style factors were assumed orthogonal to the substantive affect factor (when substantive affect and response style factors were allowed to correlate, they were found to share less than 5% of the variance). All models were estimated using maximum likelihood estimation with numerical integration in Mplus version 7.4 (Muthén & Muthén, 2015). Scale scores (for response style and affect factors) were obtained from the final models by estimating factor scores using the Expected A Posteriori method, a common scoring procedure in IRT (Falk & Cai, *in press*); these IRT scale scores were used in the subsequent analyses.

Relationships of response styles with age and cognitive abilities

Age differences in response styles were examined with regression analyses, where the response style scores were regressed on linear age; quadratic age effects were also explored. Subsequently, path analysis models were employed to test the hypothesis that cognitive abilities account for relationships between age and response styles; cognitive test scores served as intermediate variable (mediator) of the relationship between age (predictor) and response style scores (outcome) in these models. Significance tests of the mediated (indirect) effect used the product of coefficients method with bias-corrected bootstrap confidence intervals, based on 5,000 bootstrap samples (Preacher & Hayes, 2008).

Impact of correction for response styles

To evaluate whether response styles impact conclusions about age differences in positive and negative affect levels,

effects of age on affect scale scores that were either uncorrected or corrected for response styles were examined and compared. Specifically, uncorrected affect scores (obtained from IRT models without response style factor) and corrected affect scores (obtained from models that controlled for response style factors) served as simultaneous outcomes in multivariate regression analyses and were regressed on age (centered at age 50 years) and age squared. Age effects on uncorrected versus corrected affect scores were compared using Wald χ^2 tests.

Whether correction for response styles improved the convergent and predictive validity of positive and negative affect scores was tested using hierarchical multiple regression analyses. Concurrent mental health indicators (CES-D, CIDI-SF, mental health problems) and future physical health (hospital stays, onset of physical illness) served as outcomes in separate regression models, using linear regression for the (continuous) CES-D scores and logistic regressions for all other (which were all binary) health outcomes. Following recommended procedures for incremental validity testing (Haynes & Lench, 2003), uncorrected affect scores were entered into the regression first, and affect scores with response style correction were entered in a second step to test their incremental predictive effects on the health outcomes.

Results

Response Style Measurement Models

The first analysis step was to examine evidence for response styles in the positive and negative affect data. Table 2 shows the goodness of fit statistics for the nominal response IRT models compared. The AIC and BIC indices indicated that models with a single affect factor (Models N1 and P1 for negative and positive affect) had a considerably poorer fit to the data than models adding an unrestricted second response factor to the substantive affect factor (Models N2 and P2). Inspection of the category response parameters for the second factor showed that respondents high on this factor had a greater probability of choosing the extreme

response categories (“not at all” and “very much”) over other response options. This result suggested that extracting extreme response tendencies from the affect ratings improved model fit, a pattern consistent with previous findings using this modeling approach (Bolt & Johnson, 2009).

The robustness of the encountered (unrestricted) response style factor was examined in subsequent restricted models. A model in which the category response parameters of the second factor explicitly targeted extreme responses (Models N3 and P3) showed worse fit than the exploratory two-factor variant. However, fit was considerably improved when participants’ tendencies to use the lowest (“not at all”) and highest (“very much”) responses were considered as two separate style factors (Models N4 and P4); this model was more parsimonious than the unconstrained model and its fit was comparable or better (see Table 2). Factors for high and low extreme responses were positively correlated, but only moderately so ($r_s = .41$ and $.58$ for negative and positive affect items, respectively). Thus, the final model extracted two response style factors in addition to the (positive or negative) affect factor, one for the disproportionate use of low extreme (i.e., “not at all”) and one for the disproportionate use of high extreme (i.e., “very much”) response options. (The tendency to use the highest and lowest response categories could be cautiously labeled “acquiescent” and “disacquiescent” response styles. However, the more descriptive terms “high extreme” and “low extreme” response style are used here).

Response Styles in Relation to Age and Cognitive Abilities

Age differences in the (low and high) extreme response style tendencies were examined next. Linear age effects in response styles were significant, with older age predicting more low extreme (standardized $\beta = .07$, $p < .001$) and high extreme ($\beta = .04$, $p < .01$) response styles for negative affect, as well as more low ($\beta = .07$, $p < .001$) and high ($\beta = .05$, $p < .001$) extreme response styles for positive affect items. Curvilinear age effects were not significant

Table 2. Goodness of Fit Statistics for Multidimensional Nominal Response Models to Capture Affect and Response Styles

Model		Log-likelihood	BIC	AIC	Number of parameters
Negative affect					
N1	Affect factor only	-68,937.7	138,303.9	137,973.3	49
N2	Affect factor + freely estimated response style factor	-66,733.3	133,930.2	133,572.7	53
N3	Affect factor + extreme response style factor	-66,785.9	134,009.2	133,671.9	50
N4	Affect factor + low and high extreme response factors	-66,692.1	133,839.1	133,488.3	52
Positive affect					
P1	Affect factor only	-99,994.5	200,452.5	200,094.9	53
P2	Affect factor + freely estimated response style factor	-94,158.0	188,814.5	188,430.0	57
P3	Affect factor + extreme response style factor	-94,557.2	189,586.7	189,222.4	54
P4	Affect factor + low and high extreme response factors	-94,376.5	189,242.8	188,865.0	56

Notes: AIC = Akaike information criterion; BIC = Bayesian information criterion.

($ps > .13$); therefore, age was used as linear predictor in the subsequent mediation analyses.

Older age was associated with significantly lower cognitive test scores ($\beta = -.27, p < .001$). In path models controlling for age, lower cognitive abilities significantly predicted a greater tendency for low extreme response styles (negative affect: $\beta = -.06, p < .001$; positive affect: $\beta = -.15, p < .001$) and a greater tendency for high extreme response styles (negative affect: $\beta = -.11, p < .001$; positive affect: $\beta = -.17, p < .001$). (Relationships between cognitive test scores and the substantive [response style corrected] affect factors were also examined: controlling for age, higher cognitive scores significantly predicted lower negative affect [$\beta = -.16, p < .001$] and higher positive affect [$\beta = .23, p < .001$] scores, suggesting that better cognitive functioning is associated with greater affective well-being.) The indirect (mediated) effects of age via cognitive ability scores were significant in path models predicting low extreme (indirect effect = 0.016, 95% confidence interval [CI] = 0.009/0.023 for negative affect; 0.040, 95% CI = 0.033/0.049 for positive affect) and high extreme response styles (indirect effect = 0.030, 95% CI = 0.023/0.039 for negative affect; 0.047, 95% CI = 0.040/0.056 for positive affect), consistently supporting the mediation hypothesis. Controlling for cognitive ability scores, age was no longer a significant predictor of high extreme response styles ($ps > .67$), consistent with full mediation, whereas the age effects on low extreme response styles remained significant ($ps < .05$), indicating partial mediation.

Demographic differences in response styles were also observed. Being married or living together (vs. other marital status, $ps < .05$), more years of education ($ps < .001$) and White race (vs. other race, $ps < .001$) were associated with weaker tendencies for low and high extreme response styles, consistently for positive and negative affect items. Hispanics (vs. non-Hispanics, $ps < .001$) and women (vs. men, $ps < .02$) showed more extreme response tendencies for positive affect items. Results for (total and mediated) age effects did not change when these demographic covariates were controlled. (Physical health status measures—[a] number of chronic medical conditions and [b] hospital stays—reported in 2008 were also considered as potential mediators but did not significantly predict either high or low extreme response styles for positive and negative affect items [all $ps > .05$].)

Additional sensitivity analyses examined each of the four cognitive tests (immediate recall, delayed recall, serial 7s, and backward counting) separately as mediator of age effects on response styles. A total of 16 mediation models (with high and low response styles in negative and positive affect as separate outcomes) were conducted. Indirect effects for the individual cognitive tests were significant ($p < .05$) for 14 of the 16 models (nonsignificant only for effects of delayed recall and backward counting as mediators of low extreme responses in negative affect items), suggesting that each cognitive test contributed information to account for the age effects in response styles.

Impact of Response Style Correction

The final set of analyses examined the extent to which correction for response styles impacted the positive and negative affect scale scores. As illustrated in the scatterplots in Figure 1, although uncorrected and corrected scale scores were sizably correlated ($r = .897$ for negative affect and $r = .887$ for positive affect), correcting the affect scores for response styles yielded far from trivial differences in the ordering of individuals, changing participants' estimated positive and negative affect levels by as much as two z-scores.

Age trends in affect levels with and without response style correction are also shown in Figure 1. Without correction, negative affect scores decreased by about 0.3 z-scores from ages 50 to 75, with a slight subsequent increase ($\beta_{age} = -.38; \beta_{age}^2 = .30, ps < .001$); positive affect scores increased from age 50 to 65, with a decline in older age ($\beta_{age} = .26; \beta_{age}^2 = -.35, ps < .001$). Correction for response styles did not yield significantly different age gradients in positive or negative affect (see Figure 1; $ps > .06$ for Wald χ^2 tests comparing linear or quadratic age effects on uncorrected vs. corrected scores).

Whether response style correction increased the convergent and predictive validity of the affect measures was examined with hierarchical multiple regressions. Table 3

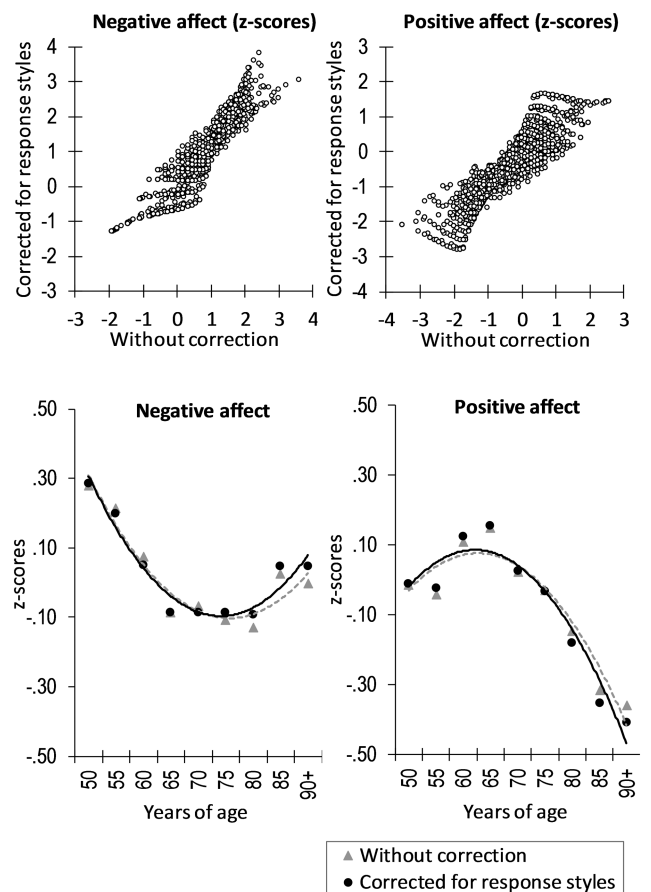


Figure 1. Scatterplots of uncorrected against response style corrected affect scale scores (top), and age trajectories in positive and negative affect based on uncorrected versus corrected scale scores (bottom).

Table 3. Hierarchical Multiple Regressions Predicting Health Outcomes From Uncorrected and Response Style Corrected Affect Scale Scores

Regression model	Concurrent mental health outcomes			Prospective physical health outcomes						
	CES-D depression		CIDI-SF major depression ^a	Diagnosis of mental health problems ^a		New illness onset ^b	Hospital stay ^c			
	β	(SE)	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)		
Negative affect										
Step 1	$R^2 = .213^{***}$		$R^2_{\text{pseudo}} = .172^{***}$		$R^2_{\text{pseudo}} = .154^{***}$		$R^2_{\text{pseudo}} = .0004$		$R^2_{\text{pseudo}} = .003^{***}$	
Uncorrected NA	.46 ^{***}	(0.02)	3.76 ^{***}	(3.20/4.43)	2.53 ^{***}	(2.33/2.75)	1.04	(0.98/1.10)	1.11 ^{***}	(1.05/1.17)
Step 2	$\Delta R^2 = .047^{***}$		$\Delta R^2_{\text{pseudo}} = .011^{***}$		$\Delta R^2_{\text{pseudo}} = .008^{***}$		$\Delta R^2_{\text{pseudo}} = .0012^*$		$\Delta R^2_{\text{pseudo}} = .003^{***}$	
Uncorrected NA	.02	(0.02)	1.49	(0.99/2.24)	1.55 ^{***}	(1.29/1.87)	0.91	(0.81/1.04)	0.9	(0.79/1.02)
Corrected NA	.49 ^{***}	(0.02)	2.08 ^{***}	(1.51/2.86)	1.58 ^{***}	(1.35/1.86)	1.15 [*]	(1.01/1.31)	1.37 ^{***}	(1.21/1.56)
Positive affect										
Step 1	$R^2 = .141^{***}$		$R^2_{\text{pseudo}} = .078^{***}$		$R^2_{\text{pseudo}} = .051^{***}$		$R^2_{\text{pseudo}} = .0014^*$		$R^2_{\text{pseudo}} = .009^{***}$	
Uncorrected PA	-.38 ^{***}	(0.02)	0.45 ^{***}	(0.40/0.52)	0.61 ^{***}	(0.57/0.66)	0.93 [*]	(0.88/0.99)	0.84 ^{***}	(0.80/0.90)
Step 2	$\Delta R^2 = .035^{***}$		$\Delta R^2_{\text{pseudo}} = .010^{***}$		$\Delta R^2_{\text{pseudo}} = .004^{***}$		$\Delta R^2_{\text{pseudo}} = .0013^*$		$\Delta R^2_{\text{pseudo}} = .005^{***}$	
Uncorrected PA	-.01	(0.03)	0.84	(0.61/1.16)	0.83 [*]	(0.70/0.97)	1.05	(0.93/1.19)	1.08	(0.96/1.21)
Corrected PA	-.40 ^{***}	(0.02)	0.55 ^{***}	(0.39/0.71)	0.73 ^{***}	(0.63/0.86)	0.88 [*]	(0.77/0.99)	0.76 ^{***}	(0.68/0.86)

Notes: CES-D = Center for Epidemiologic Studies Depression Scale; CI = confidence interval; CIDI-SF = Composite International Diagnostic Interview—Short Form; NA = negative affect; PA = positive affect; OR = odds ratios.
^aNagelkerke R^2 and ΔR^2 values are reported for dichotomous outcomes. OR are based on standardized predictor variables.
^{*} $p < .05$; ^{***} $p < .001$.

shows the results. Uncorrected affect scores (entered first into the regression) showed expected relationships with the health outcomes, with higher negative affect (and lower positive affect) predicting greater concurrent mental and prospective physical health problems (R^2 ranging from 0.04% to 21.3%, $ps < .001$; except for negative affect predicting the onset of physical illness, $p > .20$). When the response style corrected affect variables were added to the regression in the second step, they significantly augmented the prediction of each of the mental ($ps < .001$) and physical ($ps < .05$) health outcomes, both for negative (ΔR^2 ranging from 0.1% to 4.7%) and positive (ΔR^2 ranging from 0.1% to 3.5%) affect. In the final regression equations, the uncorrected affect scores did not provide a significant unique contribution to the prediction of the health outcomes (except for regressions predicting a diagnosis of mental health problems, see Table 3). Controlling for demographic covariates in the multiple regressions did not alter the pattern of results.

Discussion

Much of what we know about age-related changes in affective experiences is based on self-report questionnaires, but self-reports are sometimes fallible sources of information and potentially prone to response style effects. Using a nationally representative survey of individuals aged 50 years and older, the present results provided evidence that (a) extreme response styles in affect ratings increase in older age, (b) these age differences are at least partially accounted for by age-related declines in cognitive abilities, and (c) controlling for response styles does not impact age gradients in affect, but improves the concurrent and predictive validity of affect measures. These findings were highly consistent across positive and negative affect dimensions.

The multidimensional IRT approach used for the diagnosis of response styles detected an extreme response factor in participants' affect ratings, in line with previous research showing that extreme response styles are likely to be found in responses to rating scales (Bolt & Johnson, 2009; Cronbach, 1950; Morren et al., 2011). Importantly, the extracted response style factors capture a person's tendency to select the extreme response options independent of the person's estimated affect level, thereby isolating response patterns that do not have substantive meaning in the context of affect measurement. Distinguishing tendencies toward the lowest ("not at all") and highest ("very much") extreme category as separate factors improved model fit, potentially signifying a distinction between "acquiescent" and "disacquiescent" response styles. However, high and low extreme response styles tended to co-occur within individuals, as they were moderately positively correlated. Replicating prior research, both types of extreme responses increased similarly and in a linear fashion with participant age (Weijters et al., 2010).

Mediation analysis supported the hypothesis that increases in response styles with age are at least partly related to reductions in cognitive abilities (mental processing speed, working memory, and episodic memory) in older age. It has previously been argued that response styles may reduce the cognitive burden involved in making nuanced choices between response options (Falk & Cai, in press; Kieruj & Moors, 2013), and the results are in accordance with this. Although not directly tested here, older people may also be more easily fatigued by the sustained cognitive demands of answering a series of questionnaires—the affect questions were preceded by 26 brief scales—in that their cognitive resources may be more easily depleted by repeated survey response processes. The findings add to accumulating evidence suggesting that age-related cognitive changes have important implications for understanding how participants complete self-report ratings. For example, older people have been found to draw less on numeric values in interpreting the meaning of rating scales than younger people, which may be related to mental processing difficulties in linking the text of survey questions with the response choices (Schwarz & Knäuper, 2012). In addition, older respondents have been shown to exhibit larger response order effects (a tendency to endorse the last option in a series of response alternatives) in telephone interviews and to exhibit smaller question order effects (whereby the preceding question influences the response to the subsequent question) than younger respondents, which has been linked to age-related declines in working memory (Knäuper, Schwarz, Park, & Fritsch, 2007). These findings highlight the need for research focusing on the construction of assessment instruments that minimize the impact of age-related changes in cognitive functioning on people's responses to survey questions.

Given the significant age-related increases in response styles, a surprising result was that controlling for response styles did not noticeably alter the pattern of the age differences in positive and negative affect levels. A potential explanation is that people's tendencies toward high ("very much") and low ("not at all") extreme categories both increased with age; these tendencies "pull" (i.e., bias) the observed affect scale scores in opposite directions, and their counteracting effects may have canceled each other out. The obtained age gradients in affect replicated prior research demonstrating pronounced reductions in negative affect and stable levels of positive affect until about 70–80 years, followed by some increase in negative and decrease in positive affect in older age (Carstensen et al., 2000; Charles, Reynolds, & Gatz, 2001). Prior longitudinal studies have evidenced that this pattern is not attributable to cohort effects (Carstensen et al., 2011; Charles et al., 2001), and diary and experience sampling studies have shown that it cannot be explained by people's implicit theories or recall biases (Carstensen et al., 2000; Stone et al., 2010). The present findings add to this literature in providing evidence that age patterns in affect are not markedly distorted by age differences in response styles.

Response styles did have important implications for affect measurement in that they negatively impacted the convergent and predictive validity of the affect scale scores. Response style correction yielded affect measures with incremental validity over the uncorrected measures, significantly augmenting concurrent relationships with depression and diagnoses of mental disorders, as well as prospective associations with illness onset and hospitalizations. Even though the incremental effects were small in some instances, the consistency of results across health outcomes makes this finding especially compelling. Moreover, with one exception, the uncorrected measures did not uniquely contribute to the prediction of health outcomes beyond the corrected measures, supporting the notion that response styles dilute the construct being measured, thereby attenuating substantive effects. This underscores the importance of controlling for response styles when studying how affective experiences contribute to risks for morbidity and physiological changes accompanying the aging process.

Several study limitations should be noted. The concurrent measures of depression and mental illness were based on self-report, and these may themselves carry response style bias. In contrast to the assessment of affect, these measures were administered as structured interviews and involved dichotomous (yes/no) responses, which may reduce the complexity of the response process. However, comparisons between questionnaire and interview data can be impacted by mode effects, and studies have shown that mode effects in self-reports of depression and mental health are more pronounced with older age (Luong et al., 2015). Similarly, prospective measures of illness onset and hospital visits were also self-reported. Even though evidence suggests that people are sufficiently reliable reporters of recent medical diagnoses (El Fakiri, Bruijnzeels, & Hoes, 2007), confirmation from clinicians or medical tests would have been preferable. In addition, the cross-sectional assessment of affect and cognitive abilities precludes causal conclusions about changes in cognitive functioning and age-related increases in response styles. Future research could build on the analysis procedures used in this study to examine longitudinal changes in response styles and implications of these changes for understanding trajectories in emotional functioning across the adult life span. Finally, this study was focused on assessments of positive and negative affect and it is not clear whether the results are unique to affective experiences or generalize to other measures of psychosocial and physical functioning (many of which are assessed in the HRS). Notably, the presented multidimensional IRT model can be applied to any self-report rating scale to detect and correct for response styles. In future research, this model may prove useful to determine under which conditions (measurement content, type of response scale, survey length, etc.) age-related response style biases are most likely to occur. Pending additional research, the procedures applied here may have wide applicability for

the potential improvement of self-report measures in aging research.

Supplementary Material

Please visit the article online at <http://gerontologist.oxfordjournals.org/> to view supplementary material.

Funding

This work was supported by a grant from the National Institute on Aging (R01 AG042407).

Acknowledgments

The author is thankful to Arthur A. Stone and Doerte U. Junghaenel for their helpful comments on earlier versions of the manuscript.

References

- Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences*, *40*, 1235–1245. doi:10.1016/j.paid.2005.10.018
- Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, *7*, 608–628. doi:10.1207/S15328007SEM0704_5
- Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement*, *33*, 335–352. doi:10.1177/0146621608329891
- Bolt, D. M., Lu, Y., & Kim, J. S. (2014). Measurement and control of response styles using anchoring vignettes: A model-based approach. *Psychological Methods*, *19*, 528–541. doi:10.1037/met0000016
- Carstensen, L. L., Pasupathi, M., Mayr, U., & Nesselroade, J. R. (2000). Emotional experience in everyday life across the adult life span. *Journal of Personality and Social Psychology*, *79*, 644–655. doi:10.1037/0022-3514.79.4.644
- Carstensen, L. L., Turan, B., Scheibe, S., Ram, N., Ersner-Hershfield, H., Samanez-Larkin, G. R., ... Nesselroade, J. R. (2011). Emotional experience improves with age: Evidence based on over 10 years of experience sampling. *Psychology and Aging*, *26*, 21–33. doi:10.1037/a0021285
- Charles, S. T., & Carstensen, L. L. (2010). Social and emotional aging. *Annual Review of Psychology*, *61*, 383–409. doi:10.1146/annurev.psych.093008.100448
- Charles, S. T., Piazza, J. R., Mogle, J. A., Urban, E. J., Sliwinski, M. J., & Almeida, D. M. (in press). Age differences in emotional well-being vary by temporal recall. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*. doi:10.1093/geronb/gbv011
- Charles, S. T., Reynolds, C. A., & Gatz, M. (2001). Age-related differences and change in positive and negative affect over 23 years. *Journal of Personality and Social Psychology*, *80*, 136–151. doi:10.1037/0022-3514.80.1.136

- Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, *10*, 3–31. doi:10.1177/001316445001000101
- De Jong, M. G., Steenkamp, J.-B. E., Fox, J.-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, *45*, 104–115. doi:10.1509/jmkr.45.1.104
- Ekkkekakis, P. (2013). *The measurement of affect, mood, and emotion: A guide for health-behavioral research*. Cambridge, UK: Cambridge University Press.
- El Fakiri, F., Bruijnzeels, M. A., & Hoes, A. W. (2007). No evidence for marked ethnic differences in accuracy of self-reported diabetes, hypertension, and hypercholesterolemia. *Journal of Clinical Epidemiology*, *60*, 1271–1279. doi:10.1016/j.jclinepi.2007.02.014
- Falk, C. F., & Cai, L. (in press). A flexible full-information approach to the modeling of response styles. *Psychological Methods*. doi:10.1037/met0000059
- Fisher, G. G., Hassan, H., Faul, J. D., Rodgers, W. L., & Weir, D. R. (2015). *Health and Retirement Study imputation of cognitive functioning measures: 1992–2012*. Ann Arbor: University of Michigan.
- Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly*, *56*, 328–351. doi:10.1086/269326
- Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment*, *15*, 456–466. doi:10.1037/1040-3590.15.4.456
- Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles evidence from 19 countries. *Journal of Cross-Cultural Psychology*, *36*, 264–277. doi:10.1177/00220222022104272905
- Kieruj, N. D., & Moors, G. (2013). Response style behavior: Question format dependent or personal style? *Quality & Quantity*, *47*, 193–211. doi:10.1007/s11135-011-9511-4
- Knäuper, B., Schwarz, N., Park, D., & Fritsch, A. (2007). The perils of interpreting age differences in attitude reports: Question order effects decrease with age. *Journal of Official Statistics*, *23*, 515–528.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*, 213–236. doi:10.1002/acp.2350050305
- Luong, G., Charles, S. T., Rook, K. S., Reynolds, C. A., & Gatz, M. (2015). Age differences and longitudinal change in the effects of data collection mode on self-reports of psychosocial functioning. *Psychology and Aging*, *30*, 106–119. doi:10.1037/a0038502
- Lyrra, T.-M., Törmäkangas, T. M., Read, S., Rantanen, T., & Berg, S. (2006). Satisfaction with present life predicts survival in octogenarians. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *61*, 319–326. doi:10.1093/geronb/61.6.P319
- Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*, 209–237.
- McArdle, J. J., Fisher, G. G., & Kadlec, K. M. (2007). Latent variable analyses of age trends of cognition in the Health and Retirement Study, 1992–2004. *Psychology and Aging*, *22*, 525–545. doi:10.1037/0882-7974.22.3.525
- Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, *44*, 1539–1550. doi:10.1016/j.paid.2008.01.010
- Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & Quantity*, *42*, 779–794. doi:10.1007/s11135-006-9067-x
- Morales-Vives, F., Vigil-Colet, A., Lorenzo-Seva, U., & Ruiz-Pamies, M. (2014). How social desirability and acquiescence affects the age–personality relationship. *Personality and Individual Differences*, *60*, S16. doi:10.1016/j.paid.2013.07.370
- Morren, M., Gelissen, J. P., & Vermunt, J. K. (2011). Dealing with extreme response style in cross-cultural research: A restricted latent class factor analysis approach. *Sociological Methodology*, *41*, 13–47. doi:10.1111/j.1467-9531.2011.01238.x
- Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Nezu, A. M., Nezu, C. M., McClure, K. S., & Zwick, M. L. (2002). Assessment of depression. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression and its treatment* (pp. 61–85). New York, NY: Guilford Press.
- Ong, A. D. (2010). Pathways linking positive emotion and health in later life. *Current Directions in Psychological Science*, *19*, 358–362. doi:10.1177/0963721410388805
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightman (Eds.), *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press.
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539–569. doi:10.1146/annurev-psych-120710-100452
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891. doi:10.3758/BRM.40.3.879
- Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401. doi:10.1177/014662167700100306
- Salthouse, T. A., & Ferrer-Caja, E. (2003). What needs to be explained to account for age-related effects on multiple cognitive variables? *Psychology and Aging*, *18*, 91–110. doi:10.1037/0882-7974.18.1.91
- Schwarz, N., & Knäuper, B. (2012). Cognition, aging, and self-reports. In D. Park & N. Schwarz (Eds.), *Cognitive aging: A primer* (pp. 233–252). Philadelphia, PA: Psychology Press.
- Smith, J., Fisher, G. G., Ryan, L., Clarke, P., House, J., & Weir, D. (2013). *Health and Retirement Study Psychosocial and Lifestyle Questionnaire 2006–2010: Documentation Report*. Ann Arbor: University of Michigan.
- Soubelet, A., & Salthouse, T. A. (2011). Influence of social desirability on age differences in self-reports of mood and personality. *Journal of Personality*, *79*, 741–762. doi:10.1111/j.1467-6494.2011.00700.x
- Steffick, D. E. (2000). *Documentation of affective functioning measures in the Health and Retirement Study*. Ann Arbor: University of Michigan.
- Stephoe, A., & Wardle, J. (2011). Positive affect measured using ecological momentary assessment and survival in older men and women. *Proceedings of the National Academy of Sciences of*

- the United States of America*, 108, 18244–18248. doi:10.1073/pnas.1110892108
- Stone, A. A., Schwartz, J. E., Broderick, J. E., & Deaton, A. (2010). A snapshot of the age distribution of psychological well-being in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 9985–9990. doi:10.1073/pnas.1003744107
- Swift, H. J., Vauclair, C. M., Abrams, D., Bratt, C., Marques, S., & Lima, M. L. (2014). Revisiting the paradox of well-being: The importance of national context. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 69, 920–929. doi:10.1093/geronb/gbu011
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25, 195–217. doi:10.1093/ijpor/eds021
- Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the positive and negative affect schedule—expanded form*. University of Iowa. Unpublished manuscript.
- Weijters, B., Geuens, M., & Schillewaert, N. (2010). The stability of individual response styles. *Psychological Methods*, 15, 96–110. doi:10.1037/a0018721
- Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., & Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging*, 17, 179–193. doi:10.1037/0882-7974.17.2.179