THE GERONTOLOGICAL SOCIETY OF AMERICA®

OXFORD

Original Article

Retrospective Reports of Negative Early Life Events Over a 4-Year Period: A Test of Measurement Invariance and Response Consistency

Liat Ayalon

Louis and Gabi Weisfeld School of Social Work, Bar Ilan University, Ramat Gan, Israel.

Correspondence should be addressed to Liat Ayalon, PhD, School of Social Work, Bar Ilan University, Ramat Gan, Israel 52900; liat.ayalon@biu.ac.il

Abstract

Objectives: The present study examined measurement invariance (i.e., construct validity), response consistency (i.e., test–retest reliability), and potential predictors of response consistency to the Health and Retirement Study (HRS) negative early life events questionnaire over two time points.

Method: The study was based on the HRS psychosocial questionnaire, which is a U.S. nationally representative survey of individuals older than 50 years and their spouses of any age. Overall, 4,541 individuals older than 50 years were eligible to complete the questionnaire and responded to all four negative early life events items in 2008 and 2012.

Results: Only partial invariance across the two time points was established (with three of the four loadings and two thresholds remaining constant over time). For 20% of the sample, at least one item was inconsistently reported across waves. A positive response to a negative early life event item in 2008 was the most consistent predictor of response inconsistency over time.

Conclusions: The measure of negative early life events has limited construct validity and test–retest reliability. Inconsistency is particularly high among those who had first endorsed an item. The use of this retrospective measure for the understanding of age and aging should be considered with caution. Panel surveys might consider probing about early life events repeatedly to better address inconsistencies over time.

Keywords: Confirmatory factor analysis—Epidemiology—Life events—Psychometrics—Recall—Reliability—Retrospective—Validity

Two related theories named cumulative advantages and disadvantages attempt to explain the important role of early life events over time (Dannefer, 2003). The main notion behind these theories is that the long-term effects of early life events increase inequality between individuals (O'Rand, 1996). Negative early life events have shown to impact one's mental health (Lloyd & Turner, 2003; Shmotkin & Litwin, 2009), cognitive functioning (Brewster et al., 2014), physical health (Danese & McEwen, 2012; Grundy & Holt, 2000), and even mortality risk (Giesinger et al., 2013).

Our current knowledge about the effects of negative early life events on older adults is largely derived by large-scale epidemiological surveys such as the Health and Retirement Study (HRS), the Survey of Health, Ageing and Retirement in Europe (SHARE), and others. Because most of these large-scale epidemiological surveys sample middle-aged and older individuals and in light of the significance given to early life experiences in old age, there is a need to retrospectively assess early life experiences as part of these surveys. The life history interviews in the English Longitudinal Study of Ageing (ELSA) and SHARELIFE

(SHARE) were specifically developed to study early life events associated with late life outcomes as part of large epidemiological studies. The psychosocial questionnaire of the HRS also has a section on negative early life events. Unique to the HRS is the administration of identical sets of retrospective measures at two separate time points. This allows exploring discrepancies in retrospection across the two time points.

The present study capitalizes on the HRS repeated measurement of negative early life events to evaluate (a) measurement invariance over time, which is a form of construct validity that examines whether the same construct is being measured in the same way over time, (b) response consistency, which provides information about test–retest reliability or the degree to which test scores and items are consistent over time, and (c) potential predictors of response consistency to early life events to identify the characteristics of those individuals who are more susceptible to response inconsistency. This provides a unique contribution to current research, which has extensively relied on retrospective reports for the assessment of negative early life events in old age (Brewster et al., 2014; Logan-Greene, Green, Nurius, & Longhi, 2014), yet has failed to evaluate their psychometric properties among older adults. Given the factual nature of the negative early life events assessed (e.g., a year of school over again), response consistency is expected to be high. However, to date, the psychometric properties of the HRS negative early life events questionnaire and potential predictors of response consistency to these early life events items have not been assessed.

The HRS negative early life events questionnaire captures respondents' retrospective reports of autobiographical memories. Autobiographical memories require "a move beyond representing what happened (i.e., episodic memory) to what happened to me" (Fivush, 2011). In contrast to episodic memory, which has a defined neurological structure, autobiographical memory is only partially dependent upon neurological processes as it is formed within social, cultural, cognitive, linguistic, and communicative contexts (Nelson & Fivush, 2004). Autobiographical memories serve functional goals to provide a sense of continuity, coherence, and meaning to one's sense of self (Prebble, Addis, & Tippett, 2013) and as such, their validity and reliability highly depend upon the various contexts in which they are produced (Heidegger, 2012). For instance, early life experiences that have received only limited attention in childhood, such as the family's financial situation, are less likely to be remembered in adulthood, whereas experiences that have received considerable attention and were discussed with other family members over the years, such as the death of a parent, are more likely to be remembered (Hardt & Rutter, 2004).

Although both young and old individuals have demonstrated notable inconsistencies in their reports of autobiographical memories (Langeland et al., 2015), some have argued for an age-related decline in cognitive functioning,

in general, and autobiographic episodic memory, in particular (Habermas, Diel, & Welzer, 2013). Research has shown that compared with younger adults, older adults demonstrate a general reduction in the retrieval of autobiographical memory (McDonough & Gallo, 2013). In addition, compared with younger adults, older adults are more likely to appraise autobiographical events positively (Schryer & Ross, 2014). This has been explained by the socioemotional selectivity theory, which suggests that as older adults age and perceive their lives as limited, they are more likely to focus on the positive, rather than on negative aspects of their lives in an attempt to better regulate their mood and avoid negative emotional experiences (Mather & Carstensen, 2005).

The emotional state of the respondent also plays a role in determining reports of early life events (Dohrenwend, 2006). Individuals who have poorer mental health are more likely to report negative life events (Pachana, Brilleman, & Dobson, 2011). In support of this mood congruence theory (Bower, 1981), research has shown that individuals who experience a negative mood are particularly likely to report negative life events that are consistent with their general negative affect (Dobson, Smith, & Panchana, 2005). Similarly, individuals who suffer from posttraumatic stress disorder are more likely to report additional traumatic events (Roemer, Litz, Orsillo, Ehlich, & Friedman, 1998). Yet, others have argued that individuals who suffered from trauma are less likely to report traumatic events due to shame, attempts to repress the traumatic events, or simply forgetting of the events (Hepp et al., 2006). A different study, however, that examined errors in reports of abuse has argued that these were primarily due to unreliability in reporting or forgetfulness rather than to systematic errors due to respondents' affective state (Fergusson, Horwood, & Boden, 2011).

In addition to age and one's emotional stage, gender differences have been noted, with women being more likely to provide consistent reports of negative autobiographical events over time (Friedrich, Talley, Panser, Fett, & Zinsmeister, 1997). Ethnicity (Siddiqui, Mott, Anderson, & Flay, 1999) and education (Brener, Collins, Kann, Warren, & Williams, 1995) have also been associated with response consistency, with ethnic minorities and individuals of lower levels of education being more likely to respond inconsistently over time, compared with whites or people of higher levels of education. This has been partially attributed to limited test-taking skills (Brener et al., 1995).

The nature of the event being probed also determines its future recall probability. Research has argued that negative events (e.g., a disease of a family member) are more salient than positive ones (e.g., high grades at school; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). Yet, others have argued for exactly the opposite (Walker, Skowronski, & Thompson, 2003). Greater inconsistency has been documented with regard to less severe events, which could potentially be subject to interpretation (e.g., disease),

whereas more salient events (e.g., death) that universally are considered important are more likely to be consistently reported (Langeland et al., 2015). When questions are ambiguous (e.g., using the term laid off as a euphemism for being fired) or negatively worded, inconsistency is higher (Dillman, Smyth, & Christian, 2014; Dohrenwend, 2006). For instance, being laid off as a musician who works on a monthly contract is likely very different from the experience of being laid off as a blue-collar worker, following a factory shut down (Dohrenwend, 2006). Related to this, when different events fall within a single category, responses are more likely to be inconsistent (intracategory variability). Hence, depending on one's interpretation, being very sick can represent experiencing a severe heart attack or having the flu. Finally, more recent events are more likely to be remembered. This could partially explain the fall-off effect, namely the tendency to not report events that were previously endorsed (Pachana et al., 2011).

Although past research has noted numerous limitations associated with surveying autobiographical memories, to date, research has not specifically focused on the psychometric properties of retrospective reports of negative early life events survey items by older adults. Because of the cognitive and affective changes that take place in old age (Charles, Mather, & Carstensen, 2003; McDonough & Gallo, 2013; Schryer & Ross, 2014) and because by definition, early life events assessed in old age occurred several decades prior to the survey taking place and thus, are susceptible to inaccurate recall (Pachana et al., 2011), our current knowledge about retrospective reports of early life autobiographical memories made by adults or youth may not be generalizable.

The present study had three objectives. First, the study examined the measurement invariance of the negative early life events over two time points. Longitudinal invariance could potentially provide evidence toward construct validity (Edwards & Wirth, 2009), as one would expect a measure of negative early life events to be invariant over time (i.e., the same construct is being measured over time). Next, response consistency on the negative early life events questionnaire over the two time points was assessed (i.e., test–retest reliability). This analysis examined whether individuals provide the same responses over time and whether specific items were particularly prone to response inconsistency. Finally, potential predictors of response consistency were evaluated. This analysis identified the characteristics of certain individuals who were prone to response inconsistency. Both a direct model, in which all sociodemographic characteristics were entered simultaneously, and an indirect (mediated) model were examined.

The indirect model examined several hypotheses: (a) consistent with past research, which has shown an age-related decline in cognitive functioning (Habermas et al., 2013), it was hypothesized that as people age their cognitive functioning deteriorates, and this partially accounts for response inconsistency on the negative early life events

questions; (b) based on the socioemotional selectivity theory (Mather & Carstensen, 2005), it was hypothesized that the effects of age on response inconsistency would be partially mediated by improved mood which occurs with increasing age. In the present study, depressive symptoms assessed over two time points were used as a proxy of changes in mood.

Alternatively, one can hypothesize that depressive symptoms and cognitive functioning are not mediators, but rather effect suppressors. Because age is associated with both impaired memory and better mood, it is possible that at similar high levels of depressive symptoms, older individuals would be more inconsistent in their recall compared with younger adults. Because effect suppression and mediation are examined using the same statistical procedure, the term mediation is used throughout this article; and (c) given known gender, ethnic and educational differences in both cognitive functioning and mood (Cohen, Magai, Yaffee, & Walcott-Brown, 2005; Halpern, 2013; Kendler, Thornton, & Prescott, 2014) as well as in response consistency over time (Brener et al., 1995; Sauer, Auspurg, Hinz, & Liebig, 2011), changes in cognitive functioning and mood (e.g., depressive symptoms) were examined as potential mediators of the relationship between these demographic characteristics and response consistency over time. Figure 1 provides a pictorial account of the hypothesized model.

Method

Participants

The study is based on the psychosocial questionnaire of the HRS. The HRS is a biannual longitudinal U.S. nationally representative panel survey of individuals older than 50 years and their spouses of any age. Blacks, Latinos, and residents of Florida are oversampled. The core questionnaire collects information about income, work, assets, health, and disability over a face-to-face or a phone administration. As of 2006, the HRS has added a self-administered psychosocial questionnaire to a rotating 50% of the core panel participants. The psychosocial
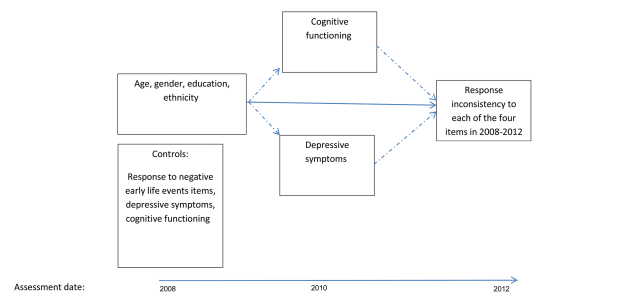
Figure 1. A conceptual model to account for response inconsistency over time. Full line represents the direct effects of independent variables on response inconsistency between 2008 and 2012. Dashed lines represent the indirect effects, which are potentially mediated through cognitive functioning and depressive symptoms assessed in 2010. All analyses were adjusted for covariates.

questionnaire evaluates life satisfaction, subjective well-being, and life circumstances. The present study is based on the 2008 and 2012 data, which represent a longitudinal panel of individuals who completed two waves of the psychosocial questionnaire. Data concerning cognitive functioning and depressive symptoms collected in 2008 and 2010 were used to model changes in cognitive functioning and mood as hypothesized by the indirect model outlined in Figure 1.

In 2008, 8,296 individuals were eligible to complete the psychosocial questionnaire. Of these individuals, 6,857 (82.6%) completed and returned the questionnaire and 100 (1.2%) completed by phone with an interviewer. Not included are 116 (1.3%) individuals who were not assigned to the psychosocial questionnaire and completed it and 1,234 (14.9%) individuals who did not return the questionnaire. In 2012, 10,079 individuals were eligible to complete the questionnaire. Of these individuals, 7,306 (72.4%) completed the questionnaire by mail, 24 (0.2%) by phone with an interviewer, 82 (0.8%) had the questionnaire completed by another person, and 2,668 (26.5%) did not return the questionnaire.

The present study concerns those individuals, older than 50 years in 2008, who were eligible to complete both waves of the psychosocial questionnaire (2008 and 2012) and provided responses to all four questions concerning negative early life events (n = 4,541). See Table 1 for characteristics of the baseline sample and differences between those who responded to all four negative early life events items in both waves and those who did not (n = 1,820). Compared with those who did not respond to both waves, those who responded to both waves were significantly younger, more educated, less depressed, of better cognitive functioning and more likely to be white.

Measures

Negative Early Life Events
As part of the psychosocial questionnaire, respondents were asked, using a yes/no response format, whether before the age of 18 (a) they had to do a year of school over again; (b) they had trouble with the police; (c) either parent drank or used drugs so that it caused problems in the family; and (d) they were physically abused by either parent (Krause, 2004). A consistent response was defined as providing the same response to the same item across the two waves (2008 and 2012; consistent = 1; inconsistent = 0).

Demographic Information
Age (in years), gender, years of education, and ethnic origin (white vs minority) were gathered based on self-report as part of the core interview.

Depressive Symptoms
The Center for Epidemiological Studies Depression scale (CES-D; Geisser, Roth, & Robinson, 1997) was used as part of the core interview to assess depressive symptoms, on eight items, using a yes–no response format. After reverse-coding appropriate items, a total score was calculated. Range was between 0 and 8, with a higher score indicating greater depressive symptoms (α_{2008} = .81; α_{2010} = .79). The eight-item CES-D has shown to be of internal consistency and factor structure which are comparable to prior versions of the CES-D (Steffick, 2000). In the present study, depressive symptoms assessed in 2008 and 2010 were used to model changes in older adults' mood.

Cognitive Functioning
A composite cognitive functioning score was constructed based on four cognitive tasks in the core questionnaire: (a) an immediate word recall task: respondents were presented with a list of 10 nouns, which they were asked to spontaneously repeat; (b) a delayed verbal memory task: after 5 minutes of engaging in other tasks, respondents were asked to repeat the list of nouns previously presented to them; (c) serial seven: respondents were asked to subtract seven from one hundred and continue subtracting seven from each subsequent number for five trials; and (d) backward count: respondents were asked to count backwards from a certain number (Fisher, Hassan, Rodgers, & Weir, 2013). A sum score was calculated with a range between 0 and 26 (a higher score indicates better cognitive functioning; α_{2008} = .65; α_{2012} = .64). In the present study, cognitive functioning assessed in 2008 and 2010 was used to model changes in cognitive functioning.

Analysis

Assessment of Measurement Invariance
To assess measurement invariance, a series of hierarchically nested confirmatory factor analysis models was examined. First, a unidimensional construct was assessed based on past research (Krause, 2004). Next, metric invariance was examined (Horn & McArdle, 1992) by setting loadings of like-indicators to be equivalent across the two waves. Finally, thresholds (given the categorical nature of indicators) of like-indicators across the two waves were set to be equivalent (i.e., scalar invariance).

Analyses were nested and sequential, with each comparison being contingent upon invariance established in previous less stringent steps of analysis. The difference between a more restricted (a model which imposes invariance) and a less restricted model was determined by comparing the fit indices. If the difference between the two models was nonsignificant, the more parsimonious model which employed invariance constrains was accepted (Vandenberg & Lance, 2000).

MPlus7 (Muthén & Muthén, 1998–2012) was used for data analyses. This program has the advantage of working with categorical indicators. Weighted least square mean variance estimation (WLSMV) was used due to the categorical nature of the indicators. The following goodness-of-fit

Table 1. Baseline Sample Characteristics and Correlations

	Incomplete data on early life events (*n* = 1,820)	Data on early life events available in both waves (*n* = 4,541)	*p* value	1	2	3	4	5	6	7	8	9
1. A year of school over again before 18 (yes)		758 (16.7%)										
2. Trouble with the police before 18 (yes)		228 (5.0%)		.10**								
3. Drinking or drugs cause family problems before 18 (yes)		735 (16.2%)		.06**	.12**							
4. Physically abused by parents before 18 (yes)		338 (7.4%)		.03	.10**	.20**						
5. Age (years), mean (*SD*)	69.3 (10.5)	68.2 (8.8)	<.001	.02	.09**	.10**	.10**					
6. Woman, frequency (percentage)	1,058 (58.1%)	2,759 (60.8%)	.05	.16**	.20**	−.02	−.05**	−.06**				
7. Education (years), mean (*SD*)	11.9 (3.8)	12.8 (2.9)	<.001	.19**	−.01	.01	.01	−.10**	−.06**			
8. White, frequency (percentage)	1,125 (61.8%)	3,568 (78.6%)	<.001	−.03	−.01	.03	.04**	−.10**	.05**	−.18**		
9. Number of depressive symptoms (0–8), median (interquartile range)	1.00 (0.00–3.00)	0.00 (0.00–2.00)	<.001	−.07**	.001	−.06**	−.10	−.06**	.11**	−.20**	.09**	
Cognitive functioning (0–26), mean (*SD*)	12.1 (4.7)	13.9 (4.1)	<.001	.20***	−.01	−.02	−.01	−.16**	.09**	.41**	−.23**	−.21**

Notes: Chi-square analyses were conducted for categorical variables, Wilcoxon tests for skewed variables, and *t*-test analyses for continuous variables with a normal distribution.

**p < .01. ***p < .001.

statistics are reported: chi-square statistic, comparative fit index (CFI), and root mean squared error (RMSEA). CFI that exceeds .95 (Hu & Bentler, 1995) and RMSEA below .08 (Musil, Jones, & Warner, 1998) are indicative of an acceptable model fit. Because WLMSV was used, DIFFTEST command and SAVADATA function in Mplus were employed to compare differences in χ^2 across nested models.

Assessment of Response Consistency

The degree of response consistency was evaluated separately for each of the items. Kappa statistics were calculated as indicators of interrater agreement (in the present

study, agreement across the two waves). Complete agreement is represented by a kappa of 1, and a kappa of 0 indicates agreement equivalent to chance. Although rules of thumb are available to estimate the size of the kappa statistics, the prevalence (i.e., the degree in which the proportion of agreement on positive classification differs from agreement on negative classification), bias (i.e., the extent to which raters disagree on the proportion of positive vs negative cases), maximum attainable kappa (i.e., the greatest possible agreement given the proportion of positive and negative values by each rater; marginal totals), and the 95% confidence interval (95% CI) of the kappa are provided to better contextualize the findings (Sim &

Wright, 2005). These analyses were conducted using Stata 13 (StataCorp, 2013).

Direct Predictors of Response Consistency

To identify predictors of response consistency, four logistic regression analyses were conducted with response consistency of each of the negative early life events as an outcome. Sociodemographic predictors, cognitive functioning, and depressive symptoms gathered in 2008 were entered simultaneously into the models. The response provided to the respective negative early life events item in 2008 was also entered into the model as a potential predictor, given the tendency of respondents to not report events which were previously endorsed (Pachana et al., 2011). This analysis was conducted using Stata 13 (StataCorp, 2013). In an additional sensitivity analyses, a composite memory score (consisting of the measures of immediate and delayed recall) was used instead of the more general measure of cognitive functioning. In addition, age was dichotomized to represent individuals older than 85 years versus individuals younger than 85 years.

Indirect Predictors of Response Consistency

A mediation analysis was conducted to identify whether changes in depressive symptoms or cognitive functioning assessed in 2008–2010 mediated the relationship between age, gender, education, or ethnicity (assessed in 2008) and response consistency over the two time periods (assessed in 2008–2012). Mediation was evaluated using a boot-strapping method with $n = 5,000$ resamples of the data to estimate the indirect effect in each sample and to obtain 95% CI (Kenny, 2013). Point estimates of indirect effects are considered significant when zero is not within the 95% CI. Mediation analysis was conducted using MPlus7 (Muthén & Muthén, 1998–2012). The advantage of using a path analysis over conventional logistic regression analyses is that the entire model proposed in Figure 1 is examined simultaneously. In an additional sensitivity analysis, a composite memory score (immediate and delayed recall) was used instead of the more general measure of cognitive functioning.

To account for the large sample size, a significant value of p less than .01 was used throughout all analyses.

Results

Assessment of Measurement Invariance

Measurement invariance was based on confirmatory factor analysis for categorical variables. The unidimensional construct of the negative early life events was established for each time period. To account for dependency over time, the two latent factors of negative early life events (assessed in 2008 and 2012) were allowed to correlate and the residual errors of respective indicators across the two waves were also allowed to correlate (e.g., the residual error of a year of school over again assessed in 2008 and in 2012). This resulted in an adequate model fit (CFI = .99, RMSEA 95%CI = 0.03[0.02–0.03], χ^2[df] = 63.3[15]). This analysis established that the pattern of loadings was invariant over the two waves (Horn & McArdle, 1992).

Next, all respective loadings across the two time periods were set to be equal. Equivalence implies that the indicators are related similarly to the latent factor (e.g., overall negative early life events) over the two measurement points and that the construct has the same meaning over time. This model resulted in an adequate fit that was not significantly worse than the unrestricted model (CFI = .99, RMSEA 95% CI = 0.02[0.02–0.03], χ^2[df] = 69.7[19]; $\Delta \chi^2$[df] = 4.3[4], p = .37).

Finally, all respective thresholds were set to be equal across time periods. This implied that the proportions of endorsement of like-indicators were equivalent across the two waves, given the same levels of the latent construct. However, this model resulted in a fit that was significantly worse than the less restricted model that assumed only loading invariance (CFI = .99, RMSEA 95% CI = .03 [0.02–0.03], χ^2[df] = 96.8[23]; $\Delta \chi^2$[df] = 47.8[4], $p < .001$).

The increasing size of the thresholds over time ($\tau_{1T1} = 1.00$, $\tau_{2T1} = 1.89$, $\tau_{3T1} = 1.33$, $\tau_{4T1} = 1.89$; $\tau_{1T2} = 1.10$, $\tau_{2T2} = 1.99$, $\tau_{3T2} = 1.39$, $\tau_{4T2} = 1.95$) suggested that over time, individuals were less likely to endorse all four items, under a comparable latent construct.

Based on the modification indices, partial invariance was tested by freeing the threshold of item 1 (a year of school over again). However, the model remained significantly worse than the metric invariance model (CFI = .99, RMSEA 95% CI = 0.03[0.02–0.03], χ^2[df] = 82.2[22]; $\Delta \chi^2$[df] = 19.9[3], $p < .001$).

To further test for partial invariance that would allow for a meaningful comparison of mean scores over time (Edwards & Wirth, 2009), the loading of one indicator was freed (drinking or drug problems in the family) and two thresholds were allowed to vary over time (a year of school over, drinking or drug problems in the family), this resulted in an adequate model fit that was nonsignificantly worse than the baseline model (CFI = .99, RMSEA 95% CI = 0.03[0.02–0.03], χ^2[df] = 75.20[20]; $\Delta \chi^2$[df] = 9.9[5], p = .08).

Assessment of Response Consistency Over Two Time Points

Table 2 describes response consistency to each of the four items across the two waves. For 20% of the sample, at least one item was inconsistently reported across the two waves. Between 27.5% (a year of school over again) and 46.9% (trouble with the police) of those who said "yes" in 2008, responded "no" to the same question in 2012. Among those who stated "no" in 2008, between 1.6% (trouble with the police) and 3.9% (drinking or drugs) stated "yes" to the same question in 2012. Overall, between 7.8% (drinking or drugs) and 3.8% (trouble with the police) of the responses

Table 2. Response Consistency to the Negative Early Life Events Questionnaire Across the Two Waves (*N* = 4,541)

	A year of school over again before 18	Trouble with the police before 18	Drinking or drugs cause family problems before 18	Physically abused by parents before 18
Yes in 2008	758 (16.7%)	228 (5.0%)	735 (16.2%)	338 (7.4%)
Yes in 2012	653 (14.4%)	191 (4.2%)	683 (15.0%)	308 (6.8%)
Yes in 2008 and No in 2012[a]	209 (27.6%)	107 (46.9%)	202 (27.5%)	123 (36.4%)
No in 2008 and Yes in 2012[b]	104 (2.7%)	70 (1.6%)	150 (3.9%)	93 (2.2%)
Overall inconsistent	313 (6.9%)	177 (3.8%)	352 (7.8%)	216 (4.7%)
Overall consistent	4,228 (93.1%)	4,364 (96.1%)	4,189 (92.2%)	4,325 (95.2%)
Prevalence	.69	.91	.69	.86
Bias	.02	.01	.01	.01
Kappa (95% CI)	.74 (.71–.77)	.56 (.49–.62)	.71 (.68–.74)	.64 (.59–.69)
Maximum possible kappa given the observations and marginal frequency	.91	.91	.96	.95
Observed kappa as the proportion of maximum possible kappa	.81	.61	.74	.67

Notes: Consistency was defined as providing the same response to like items across both waves; prevalence represents the degree in which the proportion of agreement on positive classification differs from agreement on negative classification. Bias represents the extent to which raters disagree on the proportion of positive or negative cases. 95% CI = 95% confidence interval; maximum kappa represents the greatest possible agreement given the proportion of positive and negative values by each rater (e.g., marginal totals).
[a]% out of "Yes" responses in 2008.
[b]% out of "No" responses in 2008.

were inconsistent across the two waves. In interpreting the moderate kappa statistics obtained, it is important to keep in mind the moderate-to-high prevalence (i.e., the degree in which the proportion of agreement on positive classification differs from agreement on negative classification) and low bias (i.e., the extent to which raters disagree on the proportion of positive vs negative cases) found. This is because when there is a high prevalence index or when bias is low, kappa is lower (Sim & Wright, 2005).

Direct Predictors of Responses Consistency Over Time

Table 1 outlines the baseline correlation matrix among the variables. Although most correlations were significant, all were low to moderate. To identify predictors of response consistency, four logistic regression analyses were conducted with response consistency of each of the negative early life events as an outcome. Sociodemographic predictors, cognitive functioning, and depressive symptoms gathered in 2008 were entered simultaneously into the models. The response provided to the respective negative early life events item in 2008 was also entered into the model as a potential predictor (see Table 3). Younger age, more years of education, white ethnicity, better cognitive functioning, and a "no" response to the question regarding "*a year of school over again*" were all associated with a consistent response to this item over the two waves. Women as opposed to men and individuals who said "no" to the question regarding "*trouble with the police*" were more

likely to be consistent in their response to this item. More years of education and a "no" response to the item "*drinking or drugs caused family problems*" were associated with a greater response consistency to this item. Better cognitive functioning and a "no" response to the item "*physically abused by parents*" predicted a more consistent response to this item over the two waves. In additional sensitivity analyses, a composite score of immediate and delayed recall was used as an indicator of memory functioning, instead of the more general indicator of overall cognitive functioning. In addition, age was categorized to represent individuals older than 85 years and individuals younger than 85 years. Results remained consisted and are available upon request.

Indirect Predictors of Response Consistency Over Time

Using bootstrapping technique, a mediation model was applied to test whether changes in cognitive functioning or depressive symptoms (assessed in 2008 and 2010) mediated the relationship between age, gender, education, or ethnicity (assessed in 2008) and response consistency over the two time points (assessed in 2008–2012). There was no evidence for mediation. In an additional sensitivity analysis, a composite score of immediate and delayed recall was used as an indicator of memory functioning, instead of the more general indicator of overall cognitive functioning. Results remained consisted. These results are available upon request.

Table 3. Predictors of Consistent Responses to the Negative Early Life Events Items Across the Two Waves

	Consistent response to the item: "A year of school over again before 18"	Consistent response to the item: "Trouble with the police before 18"	Consistent response to the item: "Drinking or drugs cause family problems before 18"	Consistent response to the item: "Physically abused by parents before 18"
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Age, years	0.97***(0.96–0.99)	1.02 (0.99–1.05)	1.02 (1.00–1.03)	1.01 (1.00–1.03)
Gender (men reference)	1.37 (1.04–1.08)	2.80***(1.03–4.26)	1.34 (1.05–1.71)	0.95 (0.69–1.33)
Education, years	1.09*** (1.04–1.14)	1.04 (0.97–01.12)	1.09***(1.04–1.14)	0.96 (0.91–1.02)
Ethnicity (minority group reference)	1.89*** (1.40–2.55)	0.75 (0.47–1.21)	1.06 (0.78–1.42)	0.99 (0.67–1.47)
Depressive symptoms (0–8)	0.93 (0.87–0.99)	0.95 (0.86–1.05)	0.94 (0.88–0.99)	0.91 (0.85–0.98)
Cognitive functioning (0–26)	1.09*** (1.05–1.13)	1.02 (0.97–1.07)	0.99 (0.96–1.02)	1.10*** (1.05–1.15)
A "no" response to the relevant negative life event item ("yes" reference)	1.82*** (1.70–1.95)	2.48*** (2.25–2.73)	1.73*** (1.63–1.83)	2.21*** (2.04–2.40)
–2 Log likelihood	1,648.9	952.1	2,060.0	1,287.1

Notes: Consistent response = 1; inconsistent response = 0. *p* value <.01 is used as a criterion for statistical significance. Four simultaneous logistic regression analyses were conducted.
***p < .001.

Discussion

Given the substantial interest in cumulative advantages and disadvantages as predictors of late life outcomes (DiPrete & Eirich, 2006), the growing number of studies that rely on retrospective reports for the understanding of late life outcomes, and the limited knowledge about the validity and reliability of the report of negative early life events by older adults, the present study is highly important.

An unexpected finding of the present study is that the measurement model achieved only partial invariance over time. Although, the meaning assigned to the items remained similar over time, at equal levels of the latent construct (e.g., negative early life events) individuals were less likely to endorse all four items in the second measurement point. This questions the construct validity of this measure.

As many as 20% of the sample provided an inconsistent response with regard to at least one of the four items examined over the two-time period of this study. A notable finding of the present study is the tendency to acknowledge negative early life events when first asked but to not report these events upon the second evaluation, 4 years later. For instance, as many as 46.9% of those who reported having trouble with the police before the age of 18 in 2008 denied having trouble in 2012. On the other hand, only 1.6% of those who denied having trouble in 2008 changed their response and reported trouble in 2012. A similar trend was evident for all four items. Although the questions about doing a year of school over again and alcohol or drug problems in the family resulted in "only" a little over one quarter of respondents changing their response from "yes" to "no" over the two time periods, this still represents a substantial inconsistency that should be taken into account.

Given the very clear direction of inconsistency, from endorsing the existence of an event in 2008 to its denial in

2012, inconsistency does not reflect a random error on the part of respondents, but rather a trend which goes in line with the limited ability of the scale to achieve scalar invariance (i.e., item thresholds were found to be inequivalent over time). These findings are consistent with past research exploring lifetime substance abuse among youth (Fendrich & Vaughn, 1994) and childhood sexual abuse among individuals older than 40 years (Langeland et al., 2015).

To account for the negation of previously endorsed responses, authors have suggested that over time, society has adopted a more negative approach toward drug use, which has led respondents to deny their past activity (Fendrich & Vaughn, 1994). The relatively short follow-up period (4 years) and the age of respondents (older than 50 years) in the present study make it unlikely for respondents to have become aware of the negative connotation of these events only in 2012. Hence, although it is possible that shame and stigma have motivated respondents' denial of events they had previously acknowledged, it remains unclear, why the events were acknowledged in the first place, as one would expect greater comfort and openness upon the second administration.

The findings point to specific population groups that are more likely to provide inconsistent responses over time. Consistent with past research (Sauer et al., 2011), individuals of lower levels of education were more likely to provide inconsistent responses on two of the four negative early life events items. In addition, individuals of lower levels of cognitive functioning were more likely to be inconsistent in their reports on two of the items, suggesting that forgetfulness or unintentional mix-up of events could potentially account for some response inconsistency. It also is possible that the provision of inconsistent responses by individuals of lower levels of education or lower cognitive functioning is due to poor test-taking skills that hamper their

adequate participation in research studies. Alternatively, it is possible that certain events, such as a year of school over again, which are likely more prevalent among individuals of lower levels of education or of lower cognitive functioning (Stern, 2006) are also less likely to be endorsed on subsequent assessment, reflecting the general tendency to negate responses that were previously endorsed. Moreover, compared with minorities, white respondents were more likely to provide a consistent response to the item "a year of school over again" suggesting that this item might be particularly susceptible to systematic inconsistency by certain population groups (e.g., minorities, individuals of lower levels of education, individuals of lower levels of cognitive functioning, and older individuals).

In accordance with past research (Siddiqui et al., 1999), compared with men, women were more likely to be consistent over time on the item which assessed trouble with the police. Given the fact that gender differences were found only on this item, which is more often endorsed by men (Kendler et al., 2014), it is possible that these gender differences reflect a general tendency to decline a positive response in subsequent assessment waves. Finally, in contrast to research on age differences in autobiographical memories of negative events (Boals, Hayslip Jr, & Banks, 2014), age was a predictor of response consistency on only one of the four items assessed (a year of school over again).

In addition to direct predictors of inconsistency, an indirect (mediation) model was examined. In contrast to the mediation hypotheses, changes in depressive symptoms (i.e., mood) or in cognitive functioning did not mediate the relationships between sociodemographic variables and response consistency over time. The findings do not provide support to the mood congruence hypothesis (Knight, Maines, & Robinson, 2002) nor to the hypothesis regarding cognitive changes as predictors of response inconsistency. The fact that response consistency was evaluated over a relatively long period of 4 years could potentially account for these findings. Possibly, the role of cognitive functioning or mood is enhanced when measurements occur in closer proximity.

When interpreting the findings it is important to remember that autobiographical memory is different form episodic memory and is not only a product of one's cognitive functioning (Fivush, 2011). Instead, autobiographical memory is highly context dependent and represents an attempt to provide meaning and coherence to one's sense of self (Nelson & Fivush, 2004). It is highly possible that the present study has failed to examine the various contexts that account for the production of autobiographical memories as part of a panel survey. These contexts could be at the macrolevel, such as societal changes in norms or values over time, or at the microlevel, such as implicit or explicit attempts to present oneself in a certain way. Related to this is the fact that the mechanisms that account for conscious report versus unplanned recall of retrospective events are likely quite different, but the present study did not distinguish between the two.

Despite its strengths and significance, the study has several limitations that should be addressed. First, in the HRS version, the negative early life events questionnaire contains only four highly diverse items. Whereas, being sexually abused by one's parent/s is no doubt a traumatic event, doing a year of school over again may or may not be a traumatic or even a negative event, as this highly depends on one's interpretation. Moreover, an item such as trouble with the police can be interpreted differently by different people and might be better answered using a continuous rather than a yes/no dichotomous response format. Similarly, the interpretation of repeating a year of school might be different among those who have had very few years of education compared with those with many years of education or among individuals of different cohorts. How events are recalled depends on their salience and interpretation by the individual. Hence, the response to these items is likely quite subjective. The events assessed are also highly variable with regard to their prevalence and are all negative in nature. Past research has argued that the salience, valence (e.g., positive, neutral, or negative), and prevalence of the event impact the consistency of retrospective reports (Hardt & Rutter, 2004). Asking about less ambiguous, more salient retrospective items could have resulted in more consistent responses over time. Finally, the absence of factual information about the events assessed further hampers the ability to corroborate the reports. The reliance on a measure of depressive symptoms as a proxy of mood changes (in the absence of a more appropriate scale) and the absence of a measure of autobiographical memory are additional limitations of this study.

It also is important to take into account the fact that a large number of respondents did not complete the early life event measure on both time points. The characteristics of those who did not complete the measure during the second time point also predicted response inconsistency, when both time points were completed. For instance, being a minority was associated with not responding to the second wave, but, if the measure was completed in the two measurement points, being a minority was associated with responding inconsistently. Hence, it is likely that the findings provide a conservative estimation of inconsistency and its predictors.

Because the HRS is such an influential survey with international offspring in almost every continent and generations of researchers who capitalize on the high-quality data provided by this panel survey, the use of cognitive interviews (Fisher & Geiselman, 1992) to better understand respondents' interpretations of these items as well as reasons for inconsistent responses over time is likely to provide invaluable information to the growing industry of large-scale epidemiological studies worldwide.

The findings suggest that the construct validity of the HRS negative early life events measure is limited, given its only partial measurement invariance over time. A considerable segment of the sample did not provide a consistent

response to at least one of the four items over the 4-year period, thus indicating limited test–retest reliability. Inconsistency was most consistently related to the actual response provided during the first assessment. Hence, further use of these retrospective survey questions to evaluate negative early life events among older adults should be viewed with caution, and a more critical discussion of the role of early life events in late life should take into account the potential inconsistent response style found among one fifth of the HRS sample in the present study.

Researchers who wish to explore the role of early life events over time might employ several means. First, at least some of the events, such as problems with the police or doing a year of school over again, could be corroborated against actual records. Even though obtaining these records might be time consuming, given the level of inconsistency found in the present study, such an evaluation is highly desirable. Relying on longitudinal studies that assess individuals over a longer time period might also be useful. For instance, the Wisconsin Longitudinal Study allows corroborating retrospective reports made in later life with reports made in adolescence.

In addition, asking about the same events in different waves might be useful as the different responses provided could be used for sensitivity analysis to evaluate the potential role of response inconsistency. Finally, past research has shown that unambiguous negative events, such as the death of a parent or a divorce, are more likely to be endorsed consistently compared with more ambiguous events, such as illness or separation (Hardt & Rutter, 2004). Therefore, researchers who wish to explore the role of negative early life events might wish to focus on these less ambiguous events.

Of note is the fact that most survey panels treat responses about early life events as completely valid and reliable and make no attempt for further queries. As a result, respondents are asked about early life events only once during the lifetime of the survey, making it impossible to evaluate consistency over time. It is important for panel surveys to incorporate some of the lessons learnt in the field of social or cognitive psychology that address autobiographical memory as different from episodic memory and as reflecting a variety of contextual influences in addition to factual events. This will help to sway current survey panel research from its positivistic approach to panel surveys as the source of "true factual knowledge" to a more balanced approach. Such an approach should accept the fact that even clear questions, which nonequivocally have only two possible responses, such as doing a year of school over again, could potentially result in substantial inconsistency over time.

Funding

References

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, **5**, 323. doi:10.1037//1089-2680.5.4.323

Boals, A., Hayslip, B. Jr, & Banks, J. B. (2014). Age differences in autobiographical memories of negative events. *The International Journal of Aging and Human Development*, **78**, 47–65. doi:10.2190/AG.78.1.d

Bower, G. H. (1981). Mood and memory. *American Psychologist*, **36**, 129–148. doi:10.1037/0003-066X.36.2.129

Brener, N. D., Collins, J. L., Kann, L., Warren, C. W., & Williams, B. I. (1995). Reliability of the youth risk behavior survey questionnaire. *American Journal of Epidemiology*, **141**, 575–580. doi:10.1016/S1054-139X(02)00339-7

Brewster, P. W., Melrose, R. J., Marquine, M. J., Johnson, J. K., Napoles, A., MacKay-Brandt, A., … Mungas, D. (2014). Life experience and demographic influences on cognitive function in older adults. *Neuropsychology*, **28**, 846–858. doi:10.1037/neu0000098

Charles, S. T., Mather, M., & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General*, **132**, 310–324. doi:10.1037/0096-3445.132.2.310

Cohen, C. I., Magai, C., Yaffee, R., & Walcott-Brown, L. (2005). Racial differences in syndromal and subsyndromal depression in an older urban population. *Psychiatric Services*, **56**, 1556–1563. doi:10.1176/appi.ps.56.12.1556

Danese, A., & McEwen, B. S. (2012). Adverse childhood experiences, allostasis, allostatic load, and age-related disease. *Physiology and Behavior*, **106**, 29–39. doi:10.1016/j.physbeh.2011.08.019

Dannefer, D. (2003). Cumulative advantage/disadvantage and the life course: Cross-fertilizing age and social science theory. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, **58**, S327–S337. doi:10.1093/geronb/58.6.S327

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. New Jersey, NJ: John Wiley and Sons.

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology*, **32**, 271–297. doi:10.1146/annurev.soc.32.061604.123127

Dobson, A., Smith, N., & Panchana, N. (2005). Some problems with life event lists and health outcomes. *International Journal of Behavioral Medicine*, **12**, 199–205. doi:10.1207/s15327558ijbm1203_9

Dohrenwend, B. P. (2006). Inventorying stressful life events as risk factors for psychopathology: Toward resolution of the problem of intracategory variability. *Psychological Bulletin*, **132**, 477.

Edwards, M. C., & Wirth, R. (2009). Measurement and the study of change. *Research in Human Development*, **6**, 74–96. doi:10.1080/15427600902911163

Fendrich, M., & Vaughn, C. M. (1994). Diminished lifetime substance use over time: An inquiry into differential underreporting. *Public Opinion Quarterly*, **58**, 96–123. doi:10.1086/269410

Fergusson, D. M., Horwood, L. J., & Boden, J. M. (2011). Structural equation modeling of repeated retrospective reports of childhood maltreatment. *International Journal of Methods in Psychiatric Research*, **20**, 93–104. doi:10.1002/mpr.337

Fisher, G. G., Hassan, H., Rodgers, W. L., Weir, D. R. (2013). Health and Retirement Study. *Imputation of cognitive functioning measures: 1992–2010 (Final release version). Data description.* Ann Arbor, MI: University of Michigan.

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview.* Springfield, IL: Charles C Thomas.

Fivush, R. (2011). The development of autobiographical memory. *Annual Review of Psychology*, **62**, 559–582. doi:10.1146/annurev.psych.121208.131702

Friedrich, W. N., Talley, N. J., Panser, L., Fett, S., & Zinsmeister, A. R. (1997). Concordance of reports of childhood abuse by adults. *Child Maltreatment*, **2**, 164–171. doi:10.1177/1077559597002002009

Geisser, M. E., Roth, R. S., & Robinson, M. E. (1997). Assessing depression among persons with chronic pain using the Center for Epidemiological Studies-Depression Scale and the Beck Depression Inventory: A comparative analysis. *The Clinical Journal of Pain*, **13**, 163–170. doi:10.1097/00002508-199706000-00011

Giesinger, I., Goldblatt, P., Howden-Chapman, P., Marmot, M., Kuh, D., & Brunner, E. (2013). Association of socioeconomic position with smoking and mortality: The contribution of early life circumstances in the 1946 birth cohort. *Journal of Epidemiology and Community Health*. doi:10.1136/jech-2013-203159

Grundy, E., & Holt, G. (2000). Adult life experiences and health in early old age in Great Britain. *Social Science and Medicine (1982)*, **51**, 1061–1074. doi:10.1016/S0277-9536(00)00023-X

Habermas, T., Diel, V., & Welzer, H. (2013). Lifespan trends of autobiographical remembering: Episodicity and search for meaning. *Consciousness and Cognition: An International Journal*, **22**, 1061–1073. doi:10.1016/j.concog.2013.07.010

Halpern, D. F. (2013). *Sex differences in cognitive abilities*. New York, NY: Psychology Press.

Hardt, J., & Rutter, M. (2004). Validity of adult retrospective reports of adverse childhood experiences: Review of the evidence. *Journal of Child Psychology and Psychiatry*, **45**, 260–273. doi:10.1111/j.1469-7610.2004.00218.x

Heidegger, M. (2012). *Autobiographical memory and the validity of retrospective reports*. New York, NY: Springer.

Hepp, U., Gamma, A., Milos, G., Eich, D., Ajdacic-Gross, V., Rössler, W., … Schnyder, U. (2006). Inconsistency in reporting potentially traumatic events. *The British Journal of Psychiatry*, **188**, 278–283. doi: 10.1192/bjp.bp.104.008102

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, **18**, 117–144. doi:10.1080/03610739208253916

Hu, L.-T., & Bentler, P. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural Equation Modeling. Concepts, Issues, and Applications* (pp. 76–99). London: Sage.

Kendler, K. S., Thornton, L. M., & Prescott, C. A. (2014). Gender differences in the rates of exposure to stressful life events and sensitivity to their depressogenic effects. *American Journal of Psychiatry*, **158**, 587–593. doi:10.1176/appi.ajp.158.4.587

Kenny, D. A. (2013). Mediation with dichotomous outcomes. Retrieved 16 May, 2015, from http://davidakenny.net/doc/dichmed.pdf

Knight, B. G., Maines, M. L., & Robinson, G. S. (2002). The effects of sad mood on memory in older adults: A test of the mood congruence effect. *Psychology and Aging*, **17**, 653–661. doi:10.1037/0882-7974.17.4.653

Krause, N. (2004). Lifetime trauma, emotional support, and life satisfaction among older adults. *The Gerontologist*, **44**, 615–623. doi:10.1093/geront/44.5.615

Langeland, W., Smit, J. H., Merckelbach, H., de Vries, G., Hoogendoorn, A. W., & Draijer, N. (2015). Inconsistent retrospective self-reports of childhood sexual abuse and their correlates in the general population. *Social Psychiatry and Psychiatric Epidemiology*, **50**, 603–612. doi:10.1007/s00127-014-0986-x

Lloyd, D. A., & Turner, R. J. (2003). Cumulative adversity and posttraumatic stress disorder: Evidence from a diverse community sample of young adults. *American Journal of Orthopsychiatry*, **73**, 381–391. doi:10.1037/0002-9432.73.4.381

Logan-Greene, P., Green, S., Nurius, P. S., & Longhi, D. (2014). Distinct contributions of adverse childhood experiences and resilience resources: A cohort analysis of adult physical and mental health. *Social Work in Health Care*, **53**, 776–797. doi:10.1080/00981389.2014.944251

Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences*, **9**, 496–502. doi:0.1016/j.tics.2005.08.005

McDonough, I. M., & Gallo, D. A. (2013). Impaired retrieval monitoring for past and future autobiographical events in older adults. *Psychology and Aging*, **28**, 457–466. doi:10.1037/a0032732

Musil, C. M., Jones, S. L., & Warner, C. D. (1998). Structural equation modeling and its relationship to multiple regression and factor analysis. *Research in Nursing and Health*, **21**, 271–281.

Muthén, L. K., & Muthén, B. O. (1998–2012). Mplus User's Guide (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: A social cultural developmental theory. *Psychological Review*, **111**, 486. doi:10.1037/0033-295X.111.2.486

O'Rand, A. M. (1996). The precious and the precocious: Understanding cumulative disadvantage and cumulative advantage over the life course. *The Gerontologist*, **36**, 230–238. doi:10.1093/geront/36.2.230

Pachana, N. A., Brilleman, S. L., & Dobson, A. J. (2011). Reporting of life events over time: Methodological issues in a longitudinal sample of women. *Psychological Assessment*, **23**, 277–281. doi:10.1037/a0021337

Prebble, S. C., Addis, D. R., & Tippett, L. J. (2013). Autobiographical memory and sense of self. *Psychological Bulletin*, **139**, 815.

Roemer, L., Litz, B. T., Orsillo, S. M., Ehlich, P. J., & Friedman, M. J. (1998). Increases in retrospective accounts of war-zone exposure over time: The role of PTSD symptom severity. *Journal of Traumatic Stress*, **11**, 597–605. doi:10.1023/A:1024469116047

Sauer, C., Auspurg, K., Hinz, T., & Liebig, S. (2011). The application of factorial surveys in general population samples: The effects of respondent age and education on response times and response consistency. *Survey Research Methods*, **5**, 89–102.

Schryer, E., & Ross, M. (2014). Does the age-related positivity effect in autobiographical recall reflect differences in appraisal or memory? *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, **69**, 548–556. doi:10.1093/geronb/gbt047

Shmotkin, D., & Litwin, H. (2009). Cumulative adversity and depressive symptoms among older adults in Israel: The differential

roles of self-oriented versus other-oriented events of potential trauma. *Social Psychiatry and Psychiatric Epidemiology*, **44**, 989–997. doi:10.1007/s00127-009-0020-x

Siddiqui, O., Mott, J. A., Anderson, T. L., & Flay, B. R. (1999). Characteristics of inconsistent respondents who have "ever used" drugs in a school-based sample. *Substance Use and Misuse*, **34**, 269–295. doi:10.3109/10826089909035646

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, **85**, 257–268.

StataCorp. (2013). Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.

Steffick, D. E. (2000). *Documentation of affective functioning measures in the Health and Retirement Study*. Ann Arbor, MI: HRS Health Working Group.

Stern, Y. (2006). Cognitive reserve and Alzheimer disease. *Alzheimer Disease and Associated Disorders*, **20**, 112–117.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, **3**, 4–70. doi:10.1177/109442810031002

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant—and memory helps to keep it that way! *Review of General Psychology*, **7**, 203–210. doi:10.1037/1089-2680.7.2.203