Does the CES-D measure a continuum from depression to happiness? Comparing substantive and artifactual models

Alex M. Wood [a,*], Peter J. Taylor [a], Stephen Joseph [b]

[a] School of Psychological Sciences, University of Manchester, Manchester, United Kingdom
[b] School of Sociology and Social Policy, University of Nottingham, Nottingham, United Kingdom

ABSTRACT

The Center for Epidemiologic Studies-Depression scale (CES-D) is one of the five most frequently used measures of depressive experiences. Previous research has suggested that the scale may consist of two separate factors of happiness and depression, respectively. However, recent methodological research has demonstrated that standard factor analysis cannot be used in this situation to demonstrate such factors are substantive. The substantive factor structure of the CES-D was therefore tested with two samples of younger ($N = 8857$; age range 27–35) and older ($N = 6125$; age range 64–65) people. Using a recent correction to CFA, we demonstrate that a two factor structure arises through purely artifactual reasons, and that the CES-D actually has only one substantive factor, providing evidence for a single continuum ranging from happiness to depression.

© 2010 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The Center for Epidemiologic Studies-Depression Scale (CES-D) (Radloff, 1977) is one of the five most frequently used self-report measures of depressive experiences (Santor et al., 2006) in psychological and psychiatric researches (Shaver and Brennan, 1990; McDowell and Kristjansson, 1996). The CES-D is a 20-item measure consisting of 16 negatively worded items (e.g., "I felt sad", "I felt I could not shake the blues even with help from my family or friends"; "I thought my life had been a failure") and 4 four positively worded items ("I felt happy"; "I enjoyed life"; "I felt that I was just as good as other people"; "I felt hopeful about the future"). Positive items are reverse coded so that scores have a potential range from 0 to 60 (Radloff, 1977), with higher scores indicating greater frequency of depressive experiences.

The CES-D has very strong psychometric properties, showing high convergent validity with both the Beck Depression Inventory (BDI) ($r = 0.81$) and the Zung measure of depression ($r = 0.90$), and has high accuracy in detecting depression amongst acute depressives (99% sensitivity), alcoholics (93% sensitivity, 83% specificity), and schizophrenics (93% sensitivity, 86% specificity) (Weissman et al., 1977). The measure is also particularly effective in detecting currently present major depression amongst elderly populations (Beekman et al., 1997). Uniquely amongst the common measures of depression, the CES-D is designed to measure depressive experiences in the general population, and thus conceptualizes depression as a continuum rather than as a dichotomous state (Radloff, 1977).

When studying entire populations, the continuum approach to depression measurement avoids many of the problems associated with restriction of range that occur when using clinical measures in non-clinical samples. This has lead to the CES-D becoming the measure of choice in research programs in social psychological studies and in psychiatry and epidemiology which study well-being in large scale population surveys (for example, the National Longitudinal Surveys administered by the US Bureau of Labor Statistics) (Shaver and Brennan, 1990).

Despite the widespread use of the CES-D, there are unresolved controversies surrounding the implications of including both positive and negative items in the scale. The inclusion of the positive items poses problems of interpretation. It has been suggested (Joseph and Lewis, 1998; Joseph, 2006, 2007; Joseph and Wood, in press) that a score of zero does not represent the absence of depression, but rather the presence of happiness. For a score of zero to occur, a person would have to give all of the negative items (e.g., "I felt sad") the lowest possible score ("rarely of none of the time"), and all of the positive items (e.g., "I felt happy") the highest possible score ("most or all of the time"). For such a person it would seem misleading to state that they have simply indicated an absence of depressive symptoms; such an individual has also clearly indicated the presence of happiness. Thus the CES-D, as conventionally coded, could be conceptualized as ranging from a

---

* Corresponding author. Tel.: +44 7790816407.
E-mail address: alex.wood@manchester.ac.uk (A.M. Wood).
URL: http://www.psych-sci.manchester.ac.uk/staff/alexwood (A.M. Wood).

positive pole (happiness), through a true zero-point, to a negative pole (depression) (Joseph, 2006; Santor, 2006; Joseph, 2007; Joseph and Wood, in press). While this seems a logical conclusion on the basis of the scales face validity, such a reconceptualization is controversial as it questions decades of research which has treated depression and happiness as separate phenomena and led to two distinct literatures.

Traditionally researchers have questioned the appropriateness of mixing positive and negative items (see Joseph, 2006). Radloff's (1977) original development paper showed that the positive and negative items loaded on separate factors. In a recent review, Shafer (2006) identified 28 studies performing factor analysis of the CES-D between 1977 and 2001 (total $N=22,340$) and presented meta-analytical evidence supporting separate positive and negative factors. If the positive and negative items genuinely belonged to different factors, then the normal one-factor coding would seem inappropriate. However, statistical procedures have advanced considerably since the majority of studies reviewed by Shafer were conducted. There is now significant concern about whether factors comprising entirely positive or negative items are genuinely substantive or arise purely as a result of measurement artifact (Schmitt and Stuits, 1985; Marsh, 1986, 1996; Woods, 2006).

As already noted, in the usual CES-D coding procedure, the four positive items are reverse coded. Thus potential problems arise when respondents do not fully respond to the change in item direction. At the most extreme, some respondents could not notice that certain items were reversed in content, or have developed a response set where they are rating all items with the same level of agreement. Alternatively, some respondents could show a slight bias where, for example, they were willing to strongly disagree with a negative item but only slightly agree with positive items. Such effects could even be due to "immune neglect", a phenomenon where people are up to 27% less likely to select items on the far left compared to the far right of a Likert type scale (Nicholls et al., 2006). After reverse coding, differential responding to positive and negative items would result in items which are coded in the same direction being correlated more strongly, biasing subsequent factor analysis. Two Monte Carlo studies have demonstrated that if only 10% of respondents respond carelessly to reverse coded items, then the existence of two factors would be inferred from the normal methods of both exploratory and confirmatory factor analysis (Schmitt and Stuits, 1985; Woods, 2006).

When two factors respectively contain only positive or negative items, the potential substantial importance of each factor is confounded with potential artifactual effects. This observation has lead to a growing consensus amongst methodologists that the normal methods of factor analysis cannot demonstrate the existence of two substantive factors under these conditions (Schmitt and Stuits, 1985; Marsh, 1986, 1996; Woods, 2006). Marsh (1986, 1996) suggests two methods for testing whether positive and negative factors are substantive. Traditionally, the substantive importance of positive and negative factors was demonstrated by showing that each factor had different patterns of correlations with other variables. This method is, however, dependant on appropriate variables being selected as outcomes.

As an alternative method, Marsh (1986, 1996) suggested that the different models are directly compared with confirmatory factor analysis (CFA). This approach is illustrated in Fig. 1, for a hypothetical measure with three positive and three negative items. In Model 1, positive and negative items load on a single factor. In Model 2, positive and negative items load on separate factors. In Model 3, each of the positive and negative items load on a single factor, however there are correlated errors between the positive items. Correlated errors represent an additional integrative force between the variables, in addition to the latent factor. Correlated errors would be expected if methodological bias was leading to a difference in responding to the
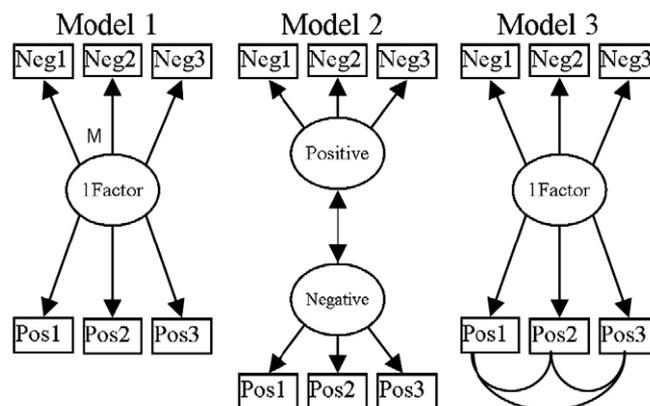

Fig. 1. Three potential models of a hypothetical measure with three positively and three negatively worded items. Model 1 has one factor, Model 2 has two factors, and Model 3 has one factor with a methodological bias.

positive items. Marsh (1986, 1996) suggests that the fit of the three models is directly compared with CFA (after adjusting for parsimony).

In this paper we examine the use two strategies to determine whether the positive and negative items form two substantively different factors, as Shafer (2006) suggests, or whether they are actually better conceptualized as one single substantive factor as assumed by the usual method of coding (Radloff, 1977; Joseph, 2006, 2007; Joseph and Wood, in press). First, we use the traditional method of testing whether the positive and negative items have different patterns of correlates, with regard to the Big Five and psychological well-being. There is now reasonable consensus that the Big Five represent most of personality at the highest level of abstraction (John and Srivastava, 1999), and that these factors are useful for orienting a scale within a map of personality psychology (Watson et al., 1994). Psychological well-being represents positive well-being, broadly defined, again at a high level of abstraction (Ryff, 1989). The choice of these correlates represents a strategy of selecting variables with breath across personality and well-being. If the positive and negative CES-D factors are substantive, then the factors would be expected to have different patterns of correlations with the Big Five and psychological well-being. Second, we more directly test whether the correlated error model (representing methodological bias) is statistically superior to a two factor model, using the approach of Marsh (1986, 1996).

2. Method

2.1. Participants and procedure

2.1.1. Older sample

The older sample comprised 6028 people aged between 64 and 65 years from the Wisconsin Longitudinal Study. This sample completed all of the variables reported in this study. These individuals are part of a long-term cohort study which follows up a group of people who graduated from high schools in Wisconsin in 1957. The current data was collected by telephone and mail interviews conducted in 2003–2005. Participants completed the CES-D and measures of the Big Five and psychological well-being. There is an approximately equal gender balance. Full demographic details and procedure are reported by Wisconsin Longitudinal Study (2006).

2.1.2. Younger sample

The younger sample comprised 8857 people aged 27–35 from the National Longitudinal Survey of Youth (NLSY). These people are part of a cohort which has been studied since 1979. This sample was included as a replication, and only completed the CES-D. The initial sampling was representative of youth in the United States, although Hispanic, Black, and economically disadvantaged youth were deliberately substantially over sampled, in order to increase generalizeability of the sample. (The sample was not weighted for oversampling as the results were not intended to be fully demographically representative, and due to the biases that can be introduced into CFA through the use of

oversampling methods.) The current data was taken from interviews in 1992. Full demographic details and procedure are reported elsewhere (Baker et al., 1993).

2.2. Measures

The *CES-D* (Radloff, 1977) is a 20-item measure of depressive experiences comprising 16 negative items (e.g., I felt sad) and four positive items (e.g., I felt happy). All questions refer to mood and attributions over the past week.

The *Big Five Inventory* (John and Srivastava, 1999) contains 44-items, with between eight and ten items measuring each of the Big Five personality traits of agreeableness, extraversion, neuroticism, conscientiousness, and openness. There is now a reasonable consensus that these traits represent most of personality at the highest level of abstraction (Watson et al., 1994). The Big Five Inventory is one of the most widely used measures of the Big Five, has internal consistency and test–retest reliability ranging from 0.79 to 0.90, and has corrected correlates with the corresponding scales of the NEO PI-R at between $r = 0.83$ and 0.99 (John and Srivastava, 1999).

The *Scales of Psychological Well-being* (Ryff, 1989) was used to assess autonomy, environmental mastery, personal growth, positive relationships, purpose in life, and self-acceptance. These variables have been designed to cover the breath of positive functioning (Ryan and Deci, 2001), and have distinct biological correlates (Ryff et al., 2006). The default coding involves six sub-scales, and this structure has recently been supported through five recent factor analyses (Ryff and Singer, 2006). However, there is a suggestion that it may be appropriate to aggregate the sub-scales of purpose in life, environmental mastery, and self-acceptance (Springer and Hauser, 2006). We used the default coding, although additionally report the results with this aggregate factor (which didn't affect the pattern of results; see the note to Table 1).

3. Results

3.1. Confirmatory factor analysis

Covariance structural equation modeling was performed using LISREL. As the items involved ordinal level responses, we used a robust weighted least squares estimation of a polychoric correlation matrix (Joreskog, 1990; Lei, 2009). Three CFA models were tested (conceptually similar to Fig. 1). A single factor model with all items loading on a single factor (Model 1), was compared with a two factor model (with positive and negative items loading on separate factors) (Model 2), and a single factor model with all items loading on a single factor, and correlated errors amongst all of the positive items (representing a method effect).

The results are shown in Table 1 for the two samples. As the models are nested, the difference in Chi squared values was used to test whether the differences in fit were statistically different. As the models had different numbers of parameters, the AIC was used to test the relative fit of the three models after adjusting for parsimony.

The results were consistent across two samples. The two factor model was superior to the one-factor model, replicating earlier studies (see Shafer, 2006). However, the correlated error model was superior to both the one and two factor models. This effect was seen across the difference in Chi square and AIC statistics. Thus, it appears that a one-factor model with a method effect best represents the data for both older and younger adults.

3.2. Correlations with the Big Five and psychological well-being

We tested whether the positive and negative items in the CES-D had different correlations with the Big Five and psychological well-being. To do this we formed two sub-scales of the CES-D, respectively comprising all positive and all negative items. The correlations (and 95% CI) are reported in Table 2. The absolute size of the correlations with outcome variables is highly similar for both the positive and negative factors. The polarity of the correlations is, however, in each case reversed. Thus the positive and negative factors respectively correlated with extroversion at $r = 0.22$ and -0.20. Although the absolute size of the correlations exceeded statistical chance for several measures (as shown by non-overlapping absolute confidence intervals), the very large sample led to very small standard errors, and differences of small substantive magnitude reached statistical significance. In each case, however, the absolute difference between the correlations is less than $r = .10$ (shared variance differed by less than 1%). These differences are small (Cohen, 1992), and the overall pattern of results shown in Table 2 involves positive and negative CES-D factors having approximately equal although reversed polarity correlations with each of the Big Five and psychological well-being variables. The approximate equality in the size of the correlations argues against the CES-D having substantively different positive and negative factors.

4. Discussion

The results suggest that the CES-D should be conceptualized as having one single underlying factor. Although a substantial number of factor analyses have been conducted on the CES-D (Shafer, 2006), all of the previous work has used standard methods of exploratory and confirmatory analyses. These studies have indicated that the CES-D has separate positive and negative factors. Were these separate factors substantive, this would bring into question the very substantial literature on depression which has used the normal single score coding of the CES-D. As the CES-D is one of the five most widely used measures of depressive experiences (Santor et al., 2006), any incorrect use of the CES-D would reduce confidence in a large body of work. However, the current research suggests that while the CES-D statistically does have separate positive and negative factors (as shown by the superiority of the two factor model to the one-factor model), these factors are not substantive and simply due to method bias (as shown by the superiority of the correlated error model to both the one and two factor models).

The study was designed to test whether positive and negative items were assessing a single factor and provide evidence that the CES-D assesses a continuum from depression to happiness. However, these results do not necessarily suggest that the CES-D does not contain separate intercorrelated factors assessing affective, motivational, somatic, and interpersonal symptoms (Shafer, 2006). The presence of these factors is of interest to psychiatric research into the architecture of depression, but investigation of these factors was not the focus of the

Table 1
Comparison of three CFA models in two samples.

Model	Model fit					Model comparisons			
	χ^2	df	AIC	CFI	RMSEA	Comparison	$\Delta\chi^2$	Δdf	ΔAIC
Older adults (N = 6028)									
1. 1 factor	2932.025	170	3012.025	0.91	0.05	1. 1 factor vs. 2 factors	830.702	1	786.702
2. 2 factors	2101.323	169	2225.323	0.94	0.04	2. Correlated errors vs. 1 factor	1094.388	6	1082.388
3. Correlated errors	1837.637	164	1929.637	0.94	0.04	3. Correlated errors vs. 2 factors	263.686	5	295.686
Younger adults (N = 8857)									
1. 1 Factor	2980.792	170	3060.792	0.91	0.04	1. 1 factor vs. 2 factors	893.512	1	849.512
2. 2 factors	2087.280	169	2211.280	0.94	0.04	2. Correlated errors vs. 1 factor	1052.175	6	1040.175
3. Correlated errors	1928.617	164	2020.617	0.95	0.04	3. Correlated errors vs. 2 factors	158.663	5	190.663

Note: All χ^2 and are $\Delta\chi^2$ are significant at $p < .001$.

Table 2
Correlations between the positive and negative factors of the CES-D and personality and well-being variables.

Variable	Correlations (95% CI)	
	Negative	Positive
Big Five		
Agreeableness	−0.20 (−0.23, −0.18)	0.22 (0.20, 0.25)
Conscientiousness	−0.28 (−0.31, 0.31)	0.21 (0.19, 0.24)
Extraversion	−0.20 (−0.23, −0.17)	0.22 (0.19, 0.25)
Neuroticism	0.42 (0.40, 0.44)	−0.33 (−0.36, −0.31)
Openness	−0.11 (−0.13, 0.08)	0.17 (0.15, 0.20)
Psychological well-being[1]		
Autonomy	−0.19 (−0.22, −0.17)	0.22 (0.20, 0.25)
Environmental mastery	−0.43 (−0.45, −0.41)	0.42 (0.40, 0.44)
Personal growth	−0.24 (−0.27, −0.22)	0.34 (0.32, 0.37)
Positive relationships	−0.34 (−0.38, −0.31)	0.35 (0.32, 0.38)
Purpose in life	−0.37 (−0.39, −0.35)	0.43 (0.42, 0.46)
Self-acceptance	−0.39 (−0.41, −0.37)	0.41 (0.39, 0.43)

*Note: N*s ranged as not all participants completed all measures. To reduce skew, measures were transformed prior to analysis. [1]The results were not changed when forming a further sub-scale by aggregating purpose in life, environmental mastery, and self-acceptance, as suggested by Springer and Hauser (2006); this scale correlated similarly with the negative (−0.44 [95% CI = −0.46, −0.43]) and positive (0.48 [95% CI = 0.46, 0.50]) CES-D sub-scales.

present work. Our research was specifically designed to ascertain whether the CES-D can continue to be scored as a single continuum despite the presence of positively and negatively worded items. Using Marsh's (1986, 1996) method, we show that seemingly separate positive and negative factors within the CES-D are simply the result of differential responding to positive and negative items. This conclusion is further supported though showing that the positive and negative latent factors have similar magnitude correlations with other variables representing the full breath of personality and positive psychological well-being.

The present results indicate that the investigation of the happiness–depression continuum is an important avenue for future research. Future research would need analysis of more representative item pools to assess the range of experiences associated with the constructs of depression and happiness. The importance of this research is clear when we consider the theoretical and practical implications of such a continuum. If the CES-D measures a continuum from depression to happiness, the theoretical question emerges about how best to conceptualize this continuum. Joseph (2006, 2007) and Joseph and Wood (in press) suggest that the constructs of depression and happiness may be largely synonymous, but existing unipolar measures simply tap different ends of the same continuum. Traditionally, we might also ask if existing research into depression now has relevance for the field of positive psychology (Linley et al., 2006). Positive psychology has a focus on happiness but not depression. But if happiness and depression were a single continuum, then studying happiness and depression separately would not make conceptual sense, and would unnecessarily duplicate research effort. One possibility is that research into depression using the CES-D may now be able to be integrated within the wider literature on the positive psychology of happiness. Similarly, if depression and happiness were on the same continuum, then this would raise the possibility that therapies that alleviate depression may also increase happiness (and vice versa) (Joseph and Wood, in press).

The present research opens up new research avenues with important theoretical and practical implications. Given the widespread use of the CES-D we recommend that researcher and practitioners using the CES-D now need to explicitly recognize that lower scores assess the presence of happiness, not simply the absence of depression.

References

Baker, C.K., Keck, F.L., Mott, S.V., Quinlan, S.V., 1993. NLSY Child Handbook, Rev. Edition: A Guide to the 1986–1990 National Longitudinal Survey of Youth Child Data. The Ohio State University, Columbus, Ohio.

Beekman, A.T.F., Deeg, D.J.H., VanLimbeek, J., Braam, A.W., DeVries, M.Z., VanTilburg, W., 1997. Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in the Netherlands. Psychological Medicine 27, 231–235.

Cohen, J., 1992. A power primer. Psychological Bulletin 112, 155–159.

John, O.P., Srivastava, S., 1999. The Big Five trait taxonomy: history, measurement, and theoretical perspectives, In: Pervin, L.A., John, O.P. (Eds.), Handbook of Personality: Theory and Research, 2nd ed. Guilford Press, New York, NY, pp. 102–138.

Joreskog, K.G., 1990. New developments in LISREL — analysis of ordinal variables using polychoric correlations and weighted least-squares. Quality & Quantity 24, 387–404.

Joseph, S., 2006. Measurement in depression: positive psychology and the statistical bipolarity of depression and happiness. Measurement 4, 156–161.

Joseph, S., 2007. Is the CES-D a measure of happiness? Psychotherapy and Psychosomatics 76, 60.

Joseph, S., Lewis, C.A., 1998. The depression–happiness scale: reliability and validity of a bipolar self report scale. Journal of Clinical Psychology 54, 537–544.

Joseph, S., Wood, A.M., in press. Assessment of positive functioning in clinical psychology: Theoretical and practical issues. Clinical Psychology Review.

Lei, P.W., 2009. Evaluating estimation methods for ordinal data in structural equation modeling. Quality & Quantity 43, 495–507.

Linley, P.A., Joseph, S., Harrington, S., Wood, A.M., 2006. Positive psychology: past, present, and (possible) future. The Journal of Positive Psychology 1, 3–16.

Marsh, H.W., 1986. Negative item bias in ratings scales for preadolescent children — a cognitive developmental phenomenon. Developmental Psychology 22, 37–49.

Marsh, H.W., 1996. Positive and negative global self-esteem: a substantively meaningful distinction or artifactors? Journal of Personality and Social Psychology 70, 810–819.

McDowell, I., Kristjansson, E., 1996. Depression. In: Newell, C. (Ed.), McDowell, I. Measuring health, A Guide to Rating Scales and Questionnaires, pp. 238–286.

Nicholls, M.E.R., Orr, C.A., Okubo, M., Loftus, A., 2006. Satisfaction guaranteed: the effect of spatial biases on responses to likert scales. Psychological Science 17, 1027–1028.

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.

Ryan, R.M., Deci, E.L., 2001. On happiness and human potentials: a review of research on hedonic and eudemonic well-being. Annual Review of Psychology 52, 141–166.

Ryff, C.D., 1989. Happiness is everything, or is it — explorations on the meaning of psychological well-being. Journal of Personality and Social Psychology 57, 1069–1081.

Ryff, C.D., Love, G.D., Urry, H.L., Muller, D., Rosenkranz, M.A., Friedman, E.M., Davidson, R.J., Singer, B., 2006. Psychological well-being and ill-being: do they have distinct or mirrored biological correlates? Psychotherapy and Psychosomatics 75, 85–95.

Ryff, C.D., Singer, B.H., 2006. Best news yet on the six-factor model of well-being. Social Science Research 35, 1103–1119.

Santor, D.A., 2006. New coke, Rosetta stones, and functional data analysis: recommendations for developing and validating new measures of depression. Measurement 4, 188–197.

Santor, D.A., Gregus, M., Welch, A., 2006. Eight decades of measurement in depression. Measurement 4, 135–155.

Schmitt, N., Stuits, D.M., 1985. Factors defined by negatively keyed items: the result of careless respondents? Applied Psychological Measurement 9, 367–373.

Shafer, A.B., 2006. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. Journal of Clinical Psychology 62, 123–146.

Shaver, P.R., Brennan, K., 1990. Measures of depression and loneliness. In: Robinson, J.P., Shaver, P.R., Wrightsman, L.S. (Eds.), Measures of Personality and Social Psychological Attitudes. Academic Press, London, pp. 195–290.

Springer, K.W., Hauser, R.M., 2006. An assessment of the construct validity of ryff's scales of psychological well-being: method, mode, and measurement effects. Social Science Research 35, 1080–1102.

Watson, D., Clark, L.A., Harkness, A.R., 1994. Structures of personality and their relevance to psychopathology. Journal of Abnormal Psychology 103, 18–31.

Weissman, M.M., Sholomskas, D., Pottenger, M., Prusoff, B.A., Locke, B.Z., 1977. Assessing depressive symptoms in five psychiatric populations: a validation study. American Journal of Epidemiology 106, 203–214.

Wisconsin Longitudinal Study, 2006. Wisconsin Longitudinal Study Handbook 2006. University of Wisconsin Survey Center, Wisconsin.

Woods, C.M., 2006. Careless responding to reverse-worded items: implications for confirmatory factor analysis. Journal of Psychopathology and Behavioral Assessment 28, 189–194.