Satisficing in Mental Healthcare Patients: The Effect of Cognitive Symptoms on Self-Report Data Quality

Judith M. Conijn,[1,2] L. Andries van der Ark,[1] Philip Spinhoven[2,3]

[1] University of Amsterdam

[2] Leiden University

[3] Leiden University Medical Center

Author Note

Judith M. Conijn, Research Institute of Child Development and Education, University of Amsterdam, The Netherlands, the Institute of Psychology, Leiden University, The Netherlands; L. Andries van der Ark, Research Institute of Child Development and Education, University of Amsterdam; Philip Spinhoven, Institute of Psychology, Leiden University, and Department of Psychiatry, Leiden University Medical Center, The Netherlands.

The infrastructure for the NESDA study (www.nesda.nl) has been funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (Zon-Mw, grant number 10-000-1002) and participating universities (VU University Medical Center, Leiden University Medical Center, University Medical Center Groningen).

Correspondence concerning this article should be addressed to Judith M. Conijn, Research Institute of Child Development and Education, University of Amsterdam, The Netherlands.

E-mail: j.m.conijn@uva.nl

Abstract

Respondents may use satisficing (i.e., nonoptimal) strategies when responding to self-report questionnaires. These satisficing strategies become more likely with decreasing motivation and/or cognitive ability (Krosnick, 1991). Considering that cognitive deficits are characteristic of depressive and anxiety disorders, depressed and anxious patients may be prone to satisficing. Using data from the Netherland's Study of Depression and Anxiety (*N* = 2,945), we studied the relationship between depression and anxiety, cognitive symptoms, and satisficing strategies on the NEO-Five Factor Inventory. Results showed that respondents with either an anxiety disorder or a comorbid anxiety and depression disorder used satisficing strategies substantially more often than healthy respondents. Cognitive symptom severity partly mediated the effect of anxiety disorder and comorbid anxiety disorder on satisficing. The results suggest that depressed and anxious patients produce relatively low quality self-report data—partly due to cognitive symptoms. Future research should investigate the degree of satisficing across different mental healthcare assessment contexts.

Keywords: careless responding, cognitive psychopathology symptoms, response inconsistency, satisficing, validity indices

Satisficing in Mental Healthcare Patients: The Effect of Cognitive Symptoms on Self-Report Data Quality

In the context of survey research, Krosnick (1991) proposed the theory of *satisficing*. Due to the cognitive effort required in responding to questionnaires, respondents with low cognitive ability or motivation may use various nonoptimal response behaviors, which Krosnick called satisficing. These satisficing strategies may vary in strength from weak satisficing, such as selecting the first alternative that seems reasonable, to strong satisficing, such as random responding. Other nonoptimal strategies include agreeing with statements regardless of content, nondifferentiation among items by repeating the same item score, or consistently selecting the “don’t know” option (Krosnick, 1991). Together, these strategies are non-content-based types of invalid responding, meaning that they are not the result of intentional deception, such as trying to make a favorable impression or achieve certain other goals.

Cognitive issues, including concentration problems, indecisiveness, memory loss, distorted thinking, and distractibility, are among the key symptoms of psychopathology and are prominent in depressive disorders (e.g., Hubbard et al., 2016). Various authors have suggested that cognitive symptoms may limit the ability to accurately complete self-report questionnaires (e.g., Cuijpers, Hofmann, & Andersson, 2010; Enns, Larsen, & Cox, 2010; Keeley, Webb, Peterson, Roussin, & Flanagan, 2016; Tada et al., 2014). Nevertheless, in large-scale studies such as the Netherland’s Study of Anxiety and Depression (NESDA; Penninx et al., 2008) or in routine outcome monitoring in clinical practice (De Beurs et al., 2011), mental healthcare patients are administered large batteries of questionnaires, which may induce satisficing strategies. On the individual patient level, satisficing may lead a clinician to under- or overestimate a patient’s symptom severity and may have negative consequences on the clinician’s decision-making process (Keeley et al., 2016). In group-level analyses, satisficing may bias research results, including observed correlations, factor structure, and group comparisons (Biderman & Reddock, 2012; Credé, 2010; Huang, Liu, & Bowling, 2015; Kam & Meyer, 2015; Osborne & Blanchard, 2011; Woods, 2006).

No previous research has explicitly assessed satisficing in mental healthcare research. In patient samples, however, different kinds of aberrant responses have been identified that may be due to satisficing, for example, “random,” “inconsistent,” or “atypical” responding (e.g., Conijn, Emons, De Jong, & Sijtsma, 2015; LePagea, Mogge, & Sharp, 2001; Wanders, Wardenaar, Penninx, Meijer, & de Jonge, 2015; Wardenaar, Wanders, Roest, Meijer, & de Jonge, 2015). In these studies, the estimated prevalence of aberrant responding ranged from 6.0% (LePagea et al., 2001) to 12.6% (Conijn et al., 2015) but cannot be directly compared due to the different detection methods used. A consistent finding is that patients with more severe psychopathology symptoms were more likely to respond aberrantly, both in nonclinical samples (Conijn, Emons, van Assen, Pedersen, & Sijtsma, 2013; Reise & Waller, 1993; Woods, Oltmanns, & Turkheimer, 2008) and clinical samples (Conijn et al., 2015; Conijn, Emons, et al. 2016; Keeley et al., 2016; Wardenaar et al., 2015). In our study, we aimed to complement previous research by addressing two limitations of previous research that are evident within the satisficing framework.

First, consistent with behavioral research (e.g., Luce, 1959; Schönberg, Daw, Joel, & O’Doherty, 2007) and experimental survey research (Mead & Craig, 2012; Peer & Gamliel, 2011), satisficing theory suggests that multiple satisficing strategies exist, including both repetitive and random strategies. However, previous research only used one type of validity indicator to assess aberrant responding among mental healthcare patients. These studies used an inconsistency scale or item response theory (IRT) based person-fit statistic (e.g., Keeley et al., 2016; Wardenaar et al., 2015). Inconsistency scales assess inconsistent responding by counting the number of inconsistent responses to highly related items (Handel, Ben-Porath, Tellegen, & Archer, 2010; Siefert et al., 2012). Person-fit statistics assess the consistency of a response pattern using the unidimensional IRT model assumed to underlie the data (Meijer, Niessen, & Tendeiro, 2016). Both inconsistency scales and person-fit statistics are effective at detecting inconsistent item scores resulting from random responding but are also sensitive enough to detect weaker forms of satisficing such as extreme response bias. However, they are unlikely to identify consistent nonoptimal response strategies, such as “don’t know” strategies or excessive utilization of the same response category. So, to comprehensively investigate satisficing in mental healthcare research, various validity indicators should be used that quantify different nonoptimal response strategies.

Second, despite the established positive relationship between psychopathology and aberrant responding (e.g., Conijn et al., 2015; Keeley et al., 2016; Wardenaar et al., 2015), the underlying explanation has not been investigated. When examining different types of disorders, a different explanation may apply. Considering depressed individuals, experimental research (Hubbard et al., 2016) combined with Krosnick’s satisficing theory provides a plausible explanation: Depressive thoughts interfere with working memory performance, resulting in problems related to concentration, (language) comprehension, and memory (Hubbard et al., 2016). In turn, these problems limit a respondent’s cognitive ability required to respond to questionnaires and likely result in a respondent employing nonoptimal response strategies (Krosnick, 1991). For respondents with a comorbid depression and anxiety disorder, the same explanation may apply because cognitive deficits have been observed to be more severe in these patients compared to patients with noncomorbid depression (e.g., Basso et al., 2007; Beaudreau & O’Hara, 2009). The relationship between anxiety disorders and cognitive impairment seems to be more complex—a possible mediating effect for cognitive symptoms is more questionable than for depression. Most studies provide evidence for poorer cognitive performance in patients with anxiety disorders or persons with high trait anxiety (Ferreri, Lapp, & Peretti, 2011; Potvin, Hudon, Dion, Grenier, & Preville, 2010; Salthouse, 2012). However, not all anxiety disorders may involve cognitive impairment (Castaneda, Tuulio-Henriksson, Marttunen, Suvisaari, & Lonnqvist, 2008), and some studies found that only patients with severe levels of anxiety show cognitive impairment, whereas those with moderately high levels of anxiety may show improved performance compared to nonanxious individuals (Bierman, Comijs, Rijmen, Jonker, & Beekman, 2008; Dotson et al., 2014).

This Study

We used Krosnick's satisficing theory to identify and explain nonoptimal response strategies in mental healthcare research. We investigated satisficing in the NESDA study, an ongoing longitudinal cohort study including five data-collection waves across a time span of 9 years. We used the baseline measurement (*n* = 2,981) that included healthy controls and participants with either a current anxiety or depression disorder or an increased risk for depressive or anxiety disorders.

Self-report questionnaires administered in NESDA include symptom scales and personality scales. We used a personality inventory, the NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992), instead of a symptom scale to investigate satisficing. Symptom scales require respondents to rate current problematic behavior (e.g., "Last week, did you worry a lot about things"), whereas personality scales require respondents to rate general behavior across a wide range of situations, including a healthy state in their past ("I'm not a worrier"). We therefore expected a personality scale to be cognitively more demanding and more relevant for studying satisficing. Satisficing was measured using multiple "satisficing indicators," coefficients that quantify a specific nonoptimal response strategy (Krosnick, 1991).

Our hypotheses were as follows:

1. Satisficing on the NEO-FFI is more common in respondents with a depression and/or anxiety disorder compared to respondents without these disorders.
2. Satisficing on the NEO-FFI is positively related to cognitive symptoms, such as problems in concentration, memory, and comprehension.
3. The severity of cognitive symptoms mediates the positive effect of having a depression and/or anxiety disorder on satisficing.

Methods

Participants and Procedure

At baseline, the NESDA study (Penninx et al., 2008) included 2,981 subjects (66% women) aged 18 to 65 years (*M* = 41.9; *SD* = 13.1). Subjects who could not speak Dutch fluently and subjects with a diagnosis of psychotic, obsessive–compulsive, bipolar, or severe addiction disorder were excluded. The baseline sample included 1,440 respondents currently diagnosed with a depression and/or anxiety disorder, 1,168 persons at risk of a depression or anxiety disorder (due to having lifetime diagnoses of depression, a family history of depression or anxiety, or subthreshold depressive or anxiety symptoms), and 373 healthy respondents. Most respondents (98%) were Dutch nationals. We excluded data from 36 respondents from our analysis due to missing scores across the complete NEO-FFI, leaving $N = 2{,}945$. In this subsample, the 918 depression diagnoses included a minor or major depressive disorder ($n = 868$) or dysthymia ($n = 275$). Anxiety disorders included social phobia ($n = 547$), panic disorder with or without agoraphobia ($n = 511$), agoraphobia ($n = 152$), and/or generalized anxiety disorder ($n = 389$)

At the baseline measurement, respondents first completed questionnaires at home (Booklet 1). One week later, trained clinical research assistants administered various observer-rated scales or interviews and experimental tasks at the research site and finally asked respondents to complete another series of questionnaires at home (Booklet 2). The NEO-FFI was the last questionnaire of Booklet 1 (pages 21 to 23). Participants were paid 15 euros for their participation and compensated for travel costs.

Measures

Depression and anxiety disorders. The lifetime version of the Composite Interview Diagnostic Instrument (CIDI; Robins et al., 1988) was used to diagnose depressive and anxiety disorders according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). The CIDI has been found to have high interrater reliability and high validity for diagnosing depressive and anxiety disorders (Wittchen, 1994).

Cognitive symptoms of psychopathology. We assessed cognitive symptoms using questions from different self-report and clinician-rated instruments concerning concentration, memory, and comprehension (Table 1). We used categorical principal components analysis (CATPCA; Linting, Meulman, Groenen, & van der Kooij, 2007) with optimal scaling in SPSS to summarize the item scores into one or several variables, while retaining maximum information from the original variable set. Inspection of eigenvalues, parallel analysis results, and component loadings showed that the data could be summarized into two correlated (r = .30) dimensions: one dimension corresponded to patient-perceived symptoms (Cronbach's α = .83) and another to clinician-perceived symptoms (Cronbach's α = .54). We concluded that dimensionality in the scores was due to mode effects (self-report vs. clinician report) instead of cognitive subdimensions (e.g., representing memory and concentration separately); therefore, we used the one-dimensional model to compute a single cognitive-symptom score, representing both the self-reported and clinician-perceived cognitive functioning. Our underlying rationale for this decision was that respondents and clinicians provide complementary information (e.g., patients provide direct insight into symptoms and a within-person comparison across time, whereas clinicians provide objective information not affected by the patient's response style or carelessness) and that their combination has the highest validity (e.g., Meyer et al., 2001). The Appendix provides more detailed results for the CATPCA—for both the two dimensional and the one-dimensional solution.

Satisficing. We assessed satisficing on the NEO-FFI, which is a shortened version of the NEO-Personality Inventory, Revised (Costa & McCrae, 1992). The NEO-FFI assesses neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. Each factor is measured using a 12-item scale, and each factor includes four to seven negatively worded items. Example items are "I'm hard-headed and tough-minded in my attitudes," "I seldom notice the moods or feelings that different environments produce," or "My life is fast-paced." Items are answered on a five-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

We computed seven satisficing indicators based on the NEO-FFI data: six response-pattern-based indices that represent five different types of nonoptimal response strategies (see Meade & Craig, 2012; Niessen, Meijer, & Tendeiro, 2016). Additionally, we used the number of missing item scores as a general satisficing indicator (Barge & Gehlbach, 2012). In the next subsections, we describe all six response-pattern-based satisficing indicators. Apart from extreme response style

(ERS) and directional response style (DRS), these satisficing behaviors are also described in Krosnick (1991).

Strong and weak nondifferentiation. We used two long string indices (DeSimone, Harms, & DeSimone, 2015) to assess consecutive repetition of responses. For every participant, we calculated the maximum length of a string of identical answers (L_{max}) and the average length of the strings of identical answers (L_{mean}). We used original item scores before recoding and ignored missing values. Furthermore, we used both of these indices to assess nondifferentiation. Researchers have found L_{max} to be somewhat more sensitive to careless responding than L_{mean} (Mead & Craig, 2012) and L_{max} may assess severe satisficing. However, L_{mean} uses all available data and may assess weaker forms of nondifferentiation compared to L_{max}.

Extreme response style. To quantify ERS, we used the percentage of valid item scores in the extreme categories. ERS is not described in Krosnick (1991) but was added based on research showing evidence for this response style (e.g., Austin, Deary, & Egan, 2006) and evidence for satisficing being an underlying cause of ERS (Aichholzer, 2013). Low motivation or low cognitive skills may result in simplifying the (Likert) response scale to a dichotomous scale with only two (extreme) options.

Midpoint response style. To quantify respondents employing “don’t know” strategies or a midpoint response style (MRS), we used the percentage of valid item scores in the middle categories.

Directional response style. Instead of the agreement response style described in Krosnick, which is the tendency to agree with statements regardless of content, we used the more general DRS, which is either the tendency to agree or disagree with statements. To quantify DRS, we subtracted the number of disagreements (< 3-score) from the number of agreements (> 3-score) and took the absolute value of the difference score. To optimally assess DRS, we only used balanced subsets of items from each NEO-FFI scale. Within the subscales, we selected items that had the highest corrected item-total correlation. This resulted in a total of 42 items. To correct for missing item scores, the DRS index was multiplied by 42 and divided by the number of valid item scores.

Random/ inconsistent responding. We used the normed version of the number of Guttman errors, also denoted as the normed *G* person-fit statistic, to detect random/inconsistent responding (Emons, 2008; Niessen et al., 2016).[1] The normed *G* statistic weighs the number of Guttman errors by the number of completed items, which varied across participants due to missing values. Because *G* normed should be applied to unidimensional data, we first assessed dimensionality of the NEO-FFI subscale data. We inspected scree plots and conducted parallel analysis using the *nFactors* package in R (Raiche, 2010). Scree plots and parallel analysis suggested unidimensionality for the Neuroticism scale, whereas the scree plot for the Extraversion and Conscientiousness scales showed unidimensionality, but parallel analysis suggested multiple factors. The Agreeableness and Openness scale showed a more substantial lack of unidimensionality. To assess whether model misfit for these subscales confounded the assessment of random/inconsistent responding with model misfit, we inspected correlations between *G* normed values computed for separate subscales. We found that the *G* normed values for the Openness and Agreeableness scales correlated equally highly with the *G* normed values for the other subscales, as the other ‘unidimensional’ subscale *G* normed values correlated with each other. We concluded that the violation of unidimensionality for the Openness and Agreeableness scales did not compromise the person-fit assessment. Subsequently, we used all NEO-FFI scales in the analysis. Using the *PerFit* package in R (Tendeiro, 2015), we computed *G* normed for every NEO-FFI subscale. Next, we averaged these values into an overall *G* normed index.

Statistical Analyses

Quantification of satisficing. We treated satisficing with respect to the NEO-FFI as a continuous variable instead of categorizing respondents into satisficers and nonsatisficers. This approach reflects that response behavior may range from using optimal strategies to using weak and strong satisficing strategies (Krosnick, 1991).[2] To assess whether we could limit the number of dependent variables in our analysis, we ran a principal component analysis (PCA) in SPSS on the six response-pattern-based satisficing indices and the number of missing responses (i.e., seven indicators in total). Using the *nFactors* package in R, we used three different methods based on the eigenvalues and scree plot to assess the number of components to retain: parallel analysis, comparing the observed eigenvalues to eigenvalues for random data; the optimal coordinate method, identifying the scree location based on the gradients associated with eigenvalues and their preceding coordinates; and the acceleration factor, which determines the coordinate where the slope of the scree plot changes most abruptly. Bartlett component scores derived from the PCA solution were used in addressing the hypotheses.

Group differences in cognitive symptom severity. We compared average cognitive symptom scores across four mutually exclusive diagnostic status categories: anxious (i.e., diagnosed with an anxiety disorder in the past month), depressed (i.e., diagnosed with a major depressive disorder and/or dysthymia in the past month), comorbid anxious and depressed, and healthy (i.e., neither depressed nor anxious in the past month).

Previous research suggests a potential nonlinear effect of anxiety on cognitive symptoms (Bierman et al., 2008; Dotson et al., 2014) and a differential effect of anxiety depending on disorder type (Canesteda et al., 2008). Therefore, we compared cognitive scores across subgroups of respondents with a different number of diagnoses (as a measure of anxiety severity) and assessed anxiety-disorder-specific effects on cognitive symptom severity. If we detected substantial nonlinear or differential effects, we took them into account in our main analyses.

Main analyses. To test whether respondents with depressive and/or anxiety disorders used satisficing strategies more frequently than respondents without these disorders (H1), we compared the mean satisficing scores across the four diagnostic status categories using multiple *t* tests. Next to comparing each of the three patient groups to the healthy group (H1), we also compared the three patient groups with each other. We used Bonferroni's correction for multiple comparisons (12 comparisons in total, six for each satisficing strategy) and Cohen's *d* to measure effect size. Cohen's *d* values of 0.2, 0.5, and 0.8 were considered indicative of small, medium, and large effects, respectively.

To test whether satisficing relates to cognitive symptoms (H2), we regressed each of the satisficing component scores on the cognitive symptom score. We used gender, age, nationality (levels: 0 = Not Dutch, 1 = Dutch), and educational level (0 = low, 1 = middle, 2 = high) as control variables in the linear regression. Next to that, we controlled for possible confounding effects of non-Dutch nationality (dummy variable) and education level, both of which may relate to satisficing through language skills and general intellectual capacity, respectively.

To test whether cognitive symptom severity mediates the effect of diagnostic status on satisficing (H3), we used the PROCESS add-on for SPSS (Hayes, 2013). We first estimated a general model in which diagnostic status was the dichotomized (0 = healthy; 1 = depression and/or anxiety disorder) independent variable, the cognitive symptom score was the mediator, and the satisficing score was the dependent variable. In this model, the control variables were the demographic variables that had a significant unique (i.e., after controlling for the other variables) relationship with the cognitive score or with satisficing. Next, we estimated similar mediation models, but now with diagnostic status as a 4-categorical independent variable. In a first type of model, we used indicator coding for diagnostic status, with the healthy group as the baseline category, and described the relationship between a specific diagnostic category (vs. being healthy) and satisficing. In the second type of model, we used sequential coding for diagnostic status to test whether an increase in satisficing in a given diagnostic patient group with respect to another diagnostic patient group was mediated by an increase in cognitive symptom severity.

In the mediation analyses, we used a stringent α level of .01 because we estimated multiple (related) mediation models. The PROCESS program uses bootstrapped confidence intervals to assess mediation effects. Mediation was assumed to occur if the 99% confidence intervals for the indirect effect (i.e., the effect of depression via cognitive symptoms on satisficing) did not contain the value 0. We assessed the size of the mediating effect by comparing the total effect of the disorder on satisficing (after accounting for the demographic control variables) to the mediating effect of the disorder on satisficing.

Finally, we conducted sensitivity analyses to assess the extent to which our conclusions regarding H2 and H3 would be altered by using the clinician-perceived cognitive symptom score derived from the two-dimensional CATPCA solution instead of the combined self- and clinician-rated score from the unidimensional solution (see Appendix).

Results

Satisficing Indicators

Descriptive statistics. Of the respondents, 10% (*n* = 298) had one to 27 missing item scores on the NEO-FFI. Most of these respondents only had one (*n* = 212), two (*n* = 52), or three (*n* = 18) missing item scores. Figure 1 shows the distribution of the six response-pattern-based satisficing indicators. For all indices, higher scores are indicative of more careless responding. Results, for example, showed that 10.8% of the respondents had a maximum long string value of at least six, 6.8% had an absolute directional bias of at least 10, 5.3% of the respondents had more than half of their responses in the extreme categories, and 8.7% had more than 40% of their responses in the middle categories.

Table 3 shows descriptive statistics and Pearson correlations for the satisficing indicators. As theoretically expected, MRS correlated negatively with ERS, and Lmax and Lmean were correlated positively. *G* normed was highly positively correlated with ERS but negatively correlated with Lmean and MRS. ERS was negatively related to both long string indices. Overall, these results suggest that repetitive responding usually does not involve the extreme categories, that random/inconsistent responding co-occurs with selecting extreme options, and that random or inconsistent responding is a different type of satisficing strategy than nondifferentiation or selecting the “don’t know” option.

Principal component analysis. All three methods for choosing the number of components to retain (parallel analysis, the optimal coordinate method, and the acceleration factor) suggested that the data were essentially two dimensional (57% variance explained). Preliminary analyses using oblique promax rotation showed that dimensions were unrelated (r = .02); therefore, we used varimax orthogonal rotation in the main analyses. The rotated factor loadings showed that the first component represented inconsistent and extreme responding, with high loadings of G normed and ERS, and was denoted as the 'erratic responding' component. The second component represented repetitive responding, with high loadings of L_{max} and L_{mean} and a moderately high loading of DRS. DRS had a substantial loading on the erratic component and a low loading on the repetitive component. The negative MRS loading on the erratic component suggested that choosing the MRS category often was a good response strategy. The negative correlations between MRS (choosing middle responses), ERS (choosing extreme responses), and G normed (choosing unexpected responses) can explain the negative MRS loading. However, the negative MRS loading was inconsistent with the underlying satisficing theory and rendered the overall assessment of satisficing as unsatisfactory. Therefore, we decided to exclude MRS from the PCA. Rerunning the PCA without MRS resulted in very similar results. Two uncorrelated dimensions adequately summarized the data (62% of the total variance explained). Table 4 shows the rotated component loadings. The main difference compared to the solution (including MRS) was that DRS now had a more substantial loading on both the erratic component and on the repetitive component. The erratic component score was skewed to the right (M = 0.0; SD = 1.0; skewness = 1.66; kurtosis = 3.95), whereas the repetitive component score was approximately normally distributed (M = 0.0; SD = 1.0; skewness = 1.27; kurtosis = 3.60).

Relationship with personality. The erratic-responding component had near-zero correlations with the NEO-FFI personality traits ($r < |.09|$). The repetitive-responding component correlated weakly with neuroticism (r = .19), extraversion (r = −.11), openness (r = −.19), and agreeableness (r = −.26). Considering the positive relationship between neuroticism and psychopathology, the pattern of correlations for repetitive responding is consistent with the expected positive relationship between depression/anxiety and satisficing.

Cognitive Symptoms

Table 2 shows descriptive statistics for the cognitive symptom score for different subgroups. The cognitive symptom score was unrelated to gender, negatively related to education level (η^2 = .03), and positively related to age (r = .04). Respondents with a non-Dutch nationality had a higher mean cognitive symptom score (Cohens' d = .38) than Dutch respondents. Compared to healthy respondents, symptom scores were substantially larger in comorbid anxious and depressed respondents (d = 1.85), in depressed respondents (d = 1.40), and in anxious respondents (d = 0.84).

For patients with anxiety disorders, we assessed whether the relationship between anxiety and cognitive symptom severity depended on the severity of anxiety (measured by the number of diagnoses) or the specific anxiety disorder. The average cognitive symptom score increased linearly with the number of diagnosed anxiety disorders—0.03 (one disorder), 0.32 (two disorders), and 0.61 (three disorders)—and was larger in each group compared to the healthy group (M = −0.54). An ANOVA showed no substantial effects of specific disorders on the cognitive score (η^2s < 0.01), after controlling for the number of anxiety diagnoses. To summarize, we found no evidence for a curvilinear effect of anxiety on cognitive symptoms or for substantial disorder-specific effects on cognitive symptoms. Therefore, we conducted the main analyses using a single anxiety category and linear effects of anxiety on the cognitive symptom score.

Main Results

Hypothesis 1. Table 5 shows the mean satisficing component scores for each diagnostic category and the effect sizes corresponding to mean-score comparisons between depressed or anxious respondents and healthy respondents. Both satisficing strategies were substantially more common in comorbid depressed and anxious respondents than in healthy respondents. Unexpectedly, depressed respondents did not show substantial mean differences in any of the satisficing scores compared to healthy respondents. Anxious patients had substantially higher scores for repetitive responding compared to healthy respondents, but showed no difference in erratic responding.

Additionally, we compared mean satisficing scores between the three patient groups. The comorbid depressed and anxious group had significantly higher mean scores on erratic responding compared to the depressed group (d = 0.28) and compared to the anxious group (d = 0.31). There were no significant group differences with respect to repetitive responding.

Hypothesis 2. The cognitive symptom score correlated .16 with inconsistent/ extreme responding and .14 with repetitive satisficing. Table 6 shows the results of the multiple regression analysis predicting satisficing scores from cognitive symptom severity and control variables. As expected, the cognitive symptom score predicted both satisficing strategies. The effect was small. The unique variance explained in satisficing by the cognitive score was 2% and 1% for erratic responding and repetitive responding, respectively. Respondents with lower education levels, non-Dutch nationality, and higher age showed more of both satisficing strategies. Gender was unrelated to satisficing.

Hypothesis 3. First, we estimated a general mediation model in which having a depression and/or anxiety disorder was the independent dummy variable, the cognitive symptom score was the mediator, and the satisficing score was the outcome variable (Figure 3). We included age, nationality, and education level as covariates. After controlling for the covariates, the total effect of the disorder dummy (see Figure 3) on erratic responding and repetitive responding was b = .21 and b = .20, respectively. Results further showed that the cognitive symptom score was a significant mediator in the relationship between depression and/or anxiety and each of the satisficing strategies. For both satisficing strategies, the indirect effect explained about half of the total effect of depression/ anxiety on satisficing (Figure 3).

Second, we assessed disorder-specific mediation effects of the cognitive symptom score on satisficing, using diagnostic status as the independent variable (Table 5, Column 1). We first compared specific disorder groups to the healthy baseline group. We only discuss the mediating effects for those disorder groups that actually had a positive mean difference in satisficing scores with respect to the healthy group (see Tables 5 and 7). For each of the three relevant comparisons, the corresponding mediating effects were significant, but effect size varied considerably (see the top rows in Table 7). The cognitive symptom score was a modest mediator in the relationship between comorbid anxiety and depression (vs. being healthy) and erratic responding. The mediating effect explained 32% of the total effect. A modest mediating effect was also found for the relationship between anxiety (vs. being healthy) and repetitive responding. In contrast, there was a large mediating effect of the cognitive score in the relationship between comorbid anxiety and depression (vs. being healthy) and repetitive responding. This effect explained 80% of the total effect.

Additional mediation analyses were conducted to assess whether the significant increase in erratic responding in the comorbid anxious and depressed group with respect to both the anxious group and the depressed group (see Table 5) was mediated by an increase in cognitive symptom severity. Both of these effects could be explained to a very small extent by a mediating effect of the cognitive symptom score (see lower rows in Table 7). In other words, differences in satisficing scores between patient groups could only be attributed to differences in cognitive symptom severity to a very small extent.

Sensitivity Analyses

We repeated the analyses using the principal component score representing the clinician-perceived cognitive problems (i.e., derived from the two-dimensional CATPCA solution; see Appendix).[3] In the multiple linear regression analyses, we found a significant but smaller effect of cognitive problems on satisficing for both erratic responding ($b = .07$; $p < .01$) and repetitive responding ($b = .11$, $p < .001$). We then re-estimated the general mediation models (see Figure 2). We could not replicate the mediating effect of cognitive symptoms in the relationship between depression and/or anxiety and erratic responding. For repetitive responding, we could confirm the mediating effect. However, the effect was smaller; the ratio between the total and direct effect equaled .14. In the disorder-specific mediation models, we could replicate three of the five mediating effects. Table 7 shows these mediating effects underlined.

Results for Midpoint Response Style

Because we excluded the MRS satisficing indicator from the PCA, we repeated the main analyses (H1–H3) using MRS as the dependent variable. ANOVA results showed that there were no significant differences in MRS between the diagnostic categories. Multiple regression analysis showed that the cognitive symptom score was significantly related to MRS after accounting for the control variables, but the effect was very small (b = .01, p < .01). We did not conduct a mediation analysis because there was no substantial relationship between having an anxiety and/or depression disorder and MRS.

Discussion

Prior research has indicated that the cognitive symptoms observed in psychopathology may interfere with valid self-report assessment (e.g., Cuijpers et al., 2010; Keeley et al., 2016; Tada et al., 2014). Furthermore, previous research has shown a relationship between cognitive ability and reporting accurately, for example, among children (Smith, Baxter, Hardin, Guinn, & Royer, 2004) and among the elderly (Wallace, Kohout, & Colsher, 1992). However, empirical support for the suggested link between cognitive symptoms and the quality of self-report data in mental healthcare patients was lacking. To investigate this relationship, we used Krosnick’s (1991) satisficing theory and chose our satisficing indicators based on recent research on the properties and performance of validity indices (Aichholzer, 2013; Mead & Craig, 2012; Niessen et al., 2016). Similar to Mead and Craig (2012), we found two dominant types of satisficing strategies: erratic (i.e., extreme or inconsistent) responding and repetitive responding.

Consistent with prior research (e.g., Keeley et al., 2016; Wardenaar et al., 2015), we found that depressed and anxious patients were more likely to satisfice on the NEO-FFI compared to healthy respondents. The effect size and type of satisficing strategy used differed across diagnostic categories. Anxious respondents used more repetitive responding compared to healthy respondents, whereas comorbid depressed and anxious respondents used both strategies more often than healthy respondents. Group differences were generally substantial but unexpectedly small when we compared depressed with healthy respondents.

Both satisficing strategies related to cognitive symptom severity. Explained variance by cognitive symptom severity was small (1–2%) but larger than the variance explained by demographic characteristics, such as education level. The low percentages of explained variance may partly be due to low reliability of the satisficing scores; however, they also suggest that variation in satisficing is largely related to other factors, such as test-taking motivation or general intelligence.

When combining disorder groups into a single patient group, results supported our hypothesis that cognitive symptom severity mediates the effects of having a depressive and/or anxiety disorder on satisficing. Further analyses of disorder-specific effects on satisficing showed that this mediating effect was only robust (or substantial) in explaining the relationships between having an anxiety disorder (with or without comorbid depression) and repetitive responding. We consider these mediating effects robust because they were also replicated using the clinician-rated cognitive score. In contrast, the mediating effect of cognitive symptom severity in the relationship between comorbid depression and anxiety and erratic responding could not be replicated using the clinician-rated score.

Considering all results, we generally found support for our three hypotheses. Patients were more likely to satisfice than healthy respondents and part of this effect was mediated by cognitive symptom severity. Concerning disorder-specific effects, we found some unexpected results; although, all results should be interpreted with caution because diagnostic specificity is limited for any diagnostic interview. Results generally suggested that other factors may also explain increased satisficing scores, especially in depressed respondents. One plausible factor represents depressive anhedonia symptoms. Anhedonic symptoms, representing lack of interest, may refer to both consummatory and motivational aspects of reward behavior. Recently, Treadway and Zald (2011) introduced the term *decisional anhedonia*, wherein the ability to balance costs and benefits when selecting among multiple options is impaired—independent from cognitive or reasoning ability. In particular this motivational and more decision-making form of anhedonia may be relevant for satisficing. Future research may assess whether or not decisional anhedonia explains additional variance in satisficing and whether or not it could provide an explanation for the low variance explained in satisficing scores in our current study.

Sensitivity analyses showed that when we used the clinician-perceived cognitive symptom score instead of the combined patient-clinician score, effect sizes were much smaller, and the mediating effect of cognitive symptom severity could only partly be replicated. These inconsistencies can be explained in several ways. A first explanation is that because the combined cognitive score reflected mainly self-reported problems, it was affected by satisficing or other response biases, such as malingering. Therefore, the regression effects in our main results were biased. Several alternative explanations relate to the quality of the clinician rating: (1) the clinical research assistants had to indirectly infer cognitive problems from a respondent's functioning during the interview; (2) research assistants could not compare the cognitive skills of patients with respect to their previous (nondepressed) functioning, so cognitive problems may not only reflect problems related to psychopathology; and (3) the rating instrument was not validated and reliability was low ($\alpha = .54$). Taken together, we can conclude that both of our alternative measures of cognitive symptoms had limitations. These limitations are strengthened by research showing a weak or nonexistent relationship between subjective (either clinician or self-report) rated cognitive performance and cognitive test performance (e.g., Homayoun, Nadeau-Marcotte, Luck, & Stip, 2010). Replication research that uses a high-quality objective measure of cognitive functioning is needed to estimate effect sizes correctly.

Our results suggest that nonoptimal response strategies may be common in mental healthcare samples. For example, we found that 10.8% of the respondents gave six identical consecutive answers at least once throughout the NEO-FFI. This response pattern is unlikely given accurate responding; the NEO-FFI items from different subscales are presented in mixed order and include positively and negatively worded items. Consider another example: 8.7% of respondents had more than 40% of their responses in the middle category, meaning that on almost half of the questions, respondents reported that they were average. It is likely that most of these respondents marked no opinion because they were satisficing. The NESDA study includes volunteers and a substantial subgroup with no current mental disorder. In other mental healthcare assessment settings (e.g., institutions where inpatients are obliged to participate in routine outcome monitoring; De Beurs et al., 2011), test-taking motivation and cognitive skills may be lower than in the NESDA sample and satisficing strategies may be more common. On the other hand, self-interest in completing questionnaires may be higher in routine practice, and the assessment may be shorter. An important topic for future research is to assess the extent to which different assessment settings induce satisficing strategies. To this end, satisficing scores on the same questionnaires could be compared between different assessment settings.

This study has several limitations. First, we did not take data on respondent motivation into account, even though lack of motivation is also a core symptom of depression. Second, we did not assess to what extent satisficing may actually be problematic in applied research using the NEO-FFI data. To what extent did satisficing bias test scores, and to what extent did that bias affect research results? In future research, these questions may be answered by excluding 5–10% of the respondents with the highest satisficing scores from the data and by assessing whether research results are substantially altered. This type of research is needed to assess the value of implementing validity indices in mental healthcare research and practice.

A third limitation is related to our approach to summarize the satisficing data. We used two dimensions of satisficing to address our hypotheses instead of the separate satisficing indicators. By using the component scores, we lost information on satisficing (18% of the total variance in the satisficing data). On the other hand, our approach probably increased the validity of the satisficing assessment. Single indicators of satisficing strategies may lack specificity. For example, prior research has suggested that person-fit statistics, such as the *G* normed statistic, may identify respondents who respond inconsistently not because they are inaccurate but because they truly have an atypical symptom or personality profile (Conrad et al., 2010; Reise & Waller, 1993; Wardenaar et al., 2015). A similar problem may apply to an index of extreme response style. Respondents may answer extremely not only because they simplify the response scale (i.e., use a satisficing strategy) but also because they are truly extreme in their behavior (e.g., He, Bartram, Inceoglu, & van de Vijver, 2014). Combining information from different validity indices may thus decrease the possibility that an unexpected response pattern is actually valid and meaningful (e.g., Conijn, Spinhoven, Meijer, & Lamers, 2016; Wanders et al., 2015).

Future research may also provide an in depth cognitive analysis of satisficing behavior in clinical samples, for example, by dichotomizing item scores into item-level satisficing indicators and by analyzing these indicators in explanatory IRT models (De Boeck & Wilson, 2004). Explanatory IRT models include person or item-explanatory variables and can be used to study which item and person properties induce satisficing strategies. Another idea for future research is to adapt decision-making models from the behavioral literature for studying satisficing strategies. For example, specific decision-making models include an autocorrelation parameter that quantifies the degree to which responses are influenced by a previous response (e.g., Lau & Glimcher, 2005; Schönberg et al., 2007). When applied to questionnaire responses, individual differences in this effect can be interpreted as differences in repetitive satisficing.

Conclusion

Our findings suggest that patients with depressive and anxiety disorders are prone to use nonoptimal response strategies on self-report measures and that cognitive symptom severity partly explains this effect. The results suggest that self-report data quality in mental healthcare research merits further attention. Future research ought to address the following questions: (1) To what extent do different healthcare assessment contexts induce satisficing strategies, (2) at what level do cognitive problems necessitate the use of rating scales instead of self-report measures, and (3) to what extent do satisficing strategies bias test scores and affect research conclusions?

Endnotes

[1] Several alternative indices can be used to assess random responding, such as person-fit statistic *lz*, or the Mahalanobis distance (e.g., Niessen et al., 2016; Mead & Craig, 2012). However, in our sample and other samples (Niessen et al. 2016) the three statistics were found to correlate highly ($r \geq .90$). Consistent with recommendations of Niessen et al. we choose the *G* person-fit statistic: (1) it imposes a less restrictive model on the data than the *lz* index, and (2) Niessen et al. found that *G* performed equally well compared to *lz* statistic but better than the Mahalanobis distance.

[2] Alternatively, satisficing may be a categorical construct, as suggested in research investigating careless responding (Meade & Craig, 2012; Kam & Meyer, 2015). Following these studies, in preliminary analyses, we used latent class profile analysis to assess whether we could identify latent satisficing groups based on the seven validity indicators. Results showed that model fit consistently improved (up to 9 classes) by adding more classes to the model, and models with better fit had a very high classification error. We concluded that a continuous quantification of satisficing would be more appropriate.

[3] The patient-perceived component score was not used in sensitivity analysis because it correlated .89 with the patient-clinician combined score.

References

Aichholzer, J. (2013). Intra-individual variation of extreme response style in mixed-mode panel Studies. *Social Science Research, 42,* 957–970. doi:10.1016/j.ssresearch.2013.01.002

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

Barge, S., & Gehlbach, G. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education, 53,* 182–200. doi:10.1007/s11162-011-9251-2

Basso, M. R., Lowery, N., Ghormley, C., Combs, D., Purdie, R., Neel, J., … Bornstein, R. (2007). Comorbid anxiety corresponds with neuropsychological dysfunction in unipolar depression. *Cognitive Neuropsychiatry, 12,* 437–456. doi:10.1080/13546800701446517

Beaudreau, S. A., & O’Hara, R. (2009). The association of anxiety and depressive symptoms with cognitive performance in community-dwelling older adults. *Psychology and Aging, 24,* 507–512. doi:10.1037/a0016035

Biderman, M. D., & Reddock, C. M. (2012). The relationship of scale reliability and validity to respondent inconsistency. *Personality and Individual Differences*, *52*, 647–651. doi:10.1016/j.paid.2011.12.012

Bierman, E. J., Comijs, H. C., Rijmen, F., Jonker, C., & Beekman, A. T. (2008) Anxiety symptoms and cognitive performance in later life: Results from the longitudinal aging study Amsterdam. *Aging and Mental Health, 12,* 517–523. doi:10.1080/13607860802224276

Buist-Bouwman, M. A., Ormel, J., De Graaf, R., Vilagut, G., Alonso, J., Van Sonderen, E., & Vollebergh, W. A. M. (2008). Psychometric properties of the World Health Organization Disability Assessment Schedule used in the European Study of the epidemiology of mental disorders. *International Journal of Methods in Psychiatric Research, 17,* 185–197. doi:10.1002/mpr.261

Castaneda, A. E., Tuulio-Henriksson, A., Marttunen, M., Suvisaari, J., & Lonnqvist, J. (2008). A review on cognitive impairments in depressive and anxiety disorders with a focus on young adults. *Journal of Affective Disorders, 106,* 1–27. doi:10.1016/j.jad.2007.06.006

Chwastiak, L. A., & Von Korff, M. (2003). Disability in depression and back pain: Evaluation of the World Health Organization Disability Assessment Schedule (WHO DAS II) in a primary care setting. *Journal of Clinical Epidemiology, 56,* 507–514. doi:10.1016/S0895-4356(03)00051-9

Conijn, J. M., Emons, W. H. M., De Jong, K., & Sijtsma, K. (2015). Detecting and explaining aberrant responding to the Outcome Questionnaire-45. *Assessment, 22,* 513–524. doi:10.1177/1073191114560882

Conijn, J. M., Emons, W. H. M., Page, B., Sijtsma, K., Van der Does, W., Carlier, I. V. E., & Giltay, E. J. (2016). Response inconsistency of patient-reported symptoms as a predictor of discrepancy between patient and clinician reported depression severity. *Assessment.* Advance online publication. doi:10.1177/1073191116666949

Conijn, J. M., Emons, W. H. M, van Assen, M. A. L. M., Pedersen, S. S., & Sijtsma, K. (2013). Explanatory, multilevel person-fit analysis of response consistency on the Spielberger State-Trait Anxiety Inventory. *Multivariate Behavioral Research, 48,* 692–718. doi:10.1080/00273171.2013.815580

Conijn, J. M., Spinhoven, P., Meijer, R. R., & Lamers, F. (2016). Person misfit on the Inventory of Depressive Symptomatology: Low quality self-report or true atypical symptom profile? *International Journal of Methods in Psychiatric Research.* Advance online publication. doi:10.1002/mpr.1548

Conrad, K. J., Bezruczko, N., Chan, Y. F., Riley, B., Diamond, G., & Dennis, M. L. (2010). Screening for atypical suicide risk with person fit statistics among people presenting to alcohol and other drug treatment. *Drug and Alcohol Dependence, 106,* 92–100. doi:10.1016/j.drugalcdep.2009.07.023

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and NEO Five-Factor Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.

Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement, 70,* 596–612. doi:10.1177/0013164410366686

Cuijpers, P., Li, J., Hofmann, S. G., & Andersson, G. (2010). Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: A meta-analysis. *Clinical Psychology Review, 30,* 768–778. doi:10.1016/j.cpr.2010.06.001

De Beurs, E., den Hollander-Gijsman, M. E., van Rood, Y. R., van der Wee, N. J. A., Giltay, E. J., van Noorden, M. S., … Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands: Practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical Psychology and Psychotherapy, 18,* 1–12. doi:10.1002/cpp.696

De Boeck, P., & Wilson, M., (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York, NY: Springer.

DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior, 36,* 171–181. doi:10.1002/job.1962

Dotson, V. M., Szymkowicz, S. M., Kirton, J. W., McLaren, M. E., Green, M., & Rohani, J. Y. (2014). Unique and interactive effect of anxiety and depressive symptoms on cognitive and brain function in young and older adults. *Journal of Depression and Anxiety, S1,* 003. doi:10.4172/2167-1044.S1-003

Emons, W. H. M. (2008). Nonparametric person-fit analysis of polytomous item scores. *Applied Psychological Measurement, 32,* 224–247. doi:10.1177/0146621607302479

Enns, M. W., Larsen, D. K., & Cox, B. J. (2000). Discrepancies between self and observer ratings of depression. The relationship to demographic, clinical and personality variables. *Journal of Affective Disorders, 60,* 33–41. doi:10.1016/S0165-0327(99)00156-1

Ferreri, F., Lapp, L. K., & Peretti, C. S. (2011). Current research on cognitive aspects of anxiety disorders. *Current Opinion in Psychiatry, 24*, 49–54. doi:10.1097/YCO.0b013e32833f5585

Handel, R. W., Ben-Porath, Y. S., Tellegen, A., & Archer, R. P. (2010). Psychometric functioning of the MMPI-2-RF VRIN-r and TRIN-r scales with varying degrees of randomness, acquiescence, and counter-acquiescence. *Psychological Assessment, 22*, 87–95. doi:10.1037/a0017061

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* New York, NY: Guilford Press.

He, J., Bartram, D., Inceoglu, I., & van de Vijver, F. J. R. (2014). Response styles and personality traits: A multilevel analysis. *Journal of Cross-Cultural Psychology*, *45*, 1028–1045. doi:10.1177/0022022114534773

Homayoun, S., Nadeau-Marcotte, F., Luck, D., & Stip, E. (2011). Subjective and objective cognitive dysfunction in schizophrenia: Is there a link? *Frontiers in Psychology, 4*(2) 148. doi:10.3389/fpsyg.2011.00148

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100,* 828–845. doi:10.1037/a0038510

Hubbard, N. A., Hutchison, J. L., Turner, M., Montroy, J., Bowles, R. P., & Rypma, B. (2016). Depressive thoughts limit working memory capacity in dysphoria. *Cognition and Emotion, 30,* 193–209. doi:10.1080/02699931.2014.991694

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods, 18,* 512–541. doi:10.1177/1094428115571894

Keeley, J. W., Webb, C., Peterson, D., Roussin, L., & Flanagan, E. H. (2016). Development of a response inconsistency scale for the personality inventory for DSM–5. *Journal of Personality Assessment*, *98,* 351–359. doi:10.1080/00223891.2016.1158719

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

Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior, 84*, 555–579. doi:10.1901/jeab.2005.110-04

LePagea, J. P., Mogge, N. L., & Sharpe, W. R. (2001). Validity rates of the MMPI-2 and PAI in a rural inpatient psychiatric facility. *Assessment, 8,* 67–74. doi:10.1177/107319110100800106

Linting, M., Meulman, J. J., Groenen, P. J. F., & van der Kooij, A. J. (2007). Nonlinear principal components analysis: Introduction and application. *Psychological Methods*, *12*, 336–358. doi:10.1037/1082-989X.12.3.336

Luce, R. (1959). *Individual choice behavior: A theoretical analysis*. New York, NY: Wiley.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17,* 437–455. doi:10.1037/a0028085

Meijer, R. R., Niessen, A. S. M., & Tendeiro, J. N. (2016). A practical guide to check the consistency of item response patterns in clinical research through person-fit statistics: Examples and a computer program. *Assessment, 23,* 52–62. doi:10.1177/1073191115577800

Meyer, G., Finn, S., Eyde, L., Kay, G., Moreland, K., Dies, R., … Reed, G. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56,* 128–165. doi:10.1037/0003-066X.56.2.128

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use? *Journal of Research in Personality, 63,* 1–11. doi:10.1016/j.jrp.2016.04.010

Osborne, J. W., & Blanchard, M. R. (2011). Random responding from participants is a threat to the validity of social science research results. *Frontiers in Psychology, 1,* 1–7. doi:10.3389/fpsyg.2010.00220

Peer, E., & Gamliel, E. (2011). Too reliable to be true? Response bias as a potential source inflation in paper-and-pencil questionnaire reliability. *Practical Assessment, Research & Evaluation, 16*(9). Retrieved from http://pareonline.net/getvn.asp?v=16&n=9

Penninx, B. W. J. H., Beekman, A. T. F., Smit, J. H., Zitman, F. G., Nolen, W. A., Spinhoven, P., … Van Dyck, R. (2008). The Netherlands study of depression and anxiety (NESDA): Rationale, objectives and methods. *International Journal of Methods in Psychiatric Research, 17*, 121–140. doi:10.1002/mpr.256.

Potvin, O., Hudon, C., Dion, M., Grenier, S., & Preville, M. (2010). Anxiety disorders, depressive episodes and cognitive impairment no dementia in community-dwelling older men and women. *International Journal of Geriatric Psychiatry, 26,* 1080–1088. doi:10.1002/gps.2647

Raiche, G. (2010). nFactors: an R package for parallel analysis and non graphical solutions to the Cattell scree test. R package version 2.3.3.

Reise, S. P., & Waller, N. G. (1993). Traitedness and the assessment of response pattern scalability. *Journal of Personality and Social Psychology, 65,* 143–151. doi:10.1037/0022-3514.65.1.143

Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F., Burke, J., … Towle, L. (1988). The composite international diagnostic interview: An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry, 45,* 1069–1077. doi:10.1001/archpsyc.1988.01800360017003

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychological Medicine, 26,* 477–486. doi:10.1017/S0033291700035558

Salthouse, T. A. (2012). How general are the effects of trait anxiety and depressive symptoms on cognitive functioning? *Emotion, 12*, 1075–1084. doi:10.1037/a0025615

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *Journal of Neuroscience, 27,* 12860–12867. doi:10.1523/JNEUROSCI.2496-07.2007

Siefert, C. J., Stein, M., Sinclair, S. J., Antonius, D., Shiva, A., & Blais, M. A. (2012). Development and initial validation of a scale for detecting inconsistent responding on the personality assessment inventory-short form. *Journal of Personality Assessment, 94,* 601–606. doi:10.1080/00223891.2012.684117

Smith, A. F, Baxter, S. D., Hardin, J. W., Guinn, C. H., & Royer, J. A. (2011). Relation of children's dietary reporting accuracy to cognitive ability. *American Journal of Epidemiology, 173,* 103–109. doi:10.1093/aje/kwq334

Jorge N. Tendeiro, Rob R. Meijer, A. Susan M. Niessen (2016). PerFit: An R package for person-fit analysis in IRT. *Journal of Statistical Software,* 74, 1-27. doi:10.18637/jss.v074.i05

Tada, M., Uchida, H., Suzuki, T., Abe, T., Pollock, B. G., & Mimura, M. (2014). Baseline difference between patients' and clinicians' rated illness severity scores and subsequent outcomes in major depressive disorder: Analysis of the sequenced treatment alternatives to relieve depression data. *Journal of Clinical Psychopharmacology, 34*, 297–302. doi:10.1097/JCP.0000000000000112

Terluin, B., van Marwijk, H. W., Ader, H. J., de Vet, H. C., Penninx, B. W., Hermens, M. L., … Stalman, W. A. (2006). The Four-Dimensional Symptom Questionnaire (4DSQ): A validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization. *BMC Psychiatry, 6,* 34. doi:10.1186/1471-244X-6-34

Treadway, M. T., & Zald, D. H. (2011). Reconsidering anhedonia in depression: Lessons from translational neuroscience. *Neuroscience and Biobehavioral Reviews, 35,* 537–555. doi:10.1016/j.neubiorev.2010.06.006

Wallace, R. B., Kohout, F. J., & Colsher, P. L. (1992). Observations on interview surveys of the oldest old. In R. M. Suzman, D. P. Willis, & K. G. Manton. (Eds.). *The oldest old* (pp. 123–134). New York, NY: Oxford University Press.

Wanders, R. B. K., Wardenaar, K. J., Penninx, B. W. J. H., Meijer, R. R., & de Jonge, P. (2015). Data-driven atypical profiles of depressive symptoms: Identification and validation in a large cohort. *Journal of Affective Disorders, 180,* 36–43. doi:10.1016/j.jad.2015.03.043

Wardenaar, K. J., Wanders, R. B. K, Roest, A. M., Meijer, R. R., & de Jonge, P. (2015). What does the beck depression inventory measure in myocardial infarction patients?: A psychometric approach using item response theory and person-fit. *International Journal of Methods in Psychiatric Research, 24,* 130–142. doi:10.1002/mpr.1467

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology, 104*, 3–14. doi:10.1037/0021-843X.104.1.3

Wittchen, H.-U. (1994). Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review. *Journal of Psychiatric Research, 28,* 57–84. doi:10.1016/0022-3956(94)90036-1

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology Behavioral Assessment, 28,* 189–194. doi:10.1007/s10862-005-9004-7

Woods, C. M., Oltmanns, T. F., & Turkheimer, E. (2008). Detection of aberrant responding on a personality scale in a military sample: An application of evaluating person fit with two-level logistic regression. *Psychological Assessment*, *20*, 159–168. doi:10.1037/1040-3590.20.2.159

Table 1

Items and Scales Used to Assess Cognitive Symptoms of Psychopathology

Scale	Item nr. / Subscale	Measuring	Mode	Scale
IDS (Rush, Gullion, Basco, Jarrett, & Trivedi, 1996)	13	Concentration and decision making	SR	4-point scale
MASQ-30 (Watson et al., 1995)	25	Difficulty in taking decisions	SR	5-point Likert
4DSQ distress subscale (Terluin et al., 2006)	12	Difficulty in thinking clearly	SR	5-point Likert
WHO-DAS-II (Chwastiak & Von Korff, 2003)	Subscale (6 items)	Communication and understanding	SR	sum score
WHODAS-II interview (last month symptoms)	5	Difficulties in concentrating, memory, and understanding things clearly	CR	yes/no
Evaluation questionnaire for the research assistant[a]	2.3	Concentration problems (during the interview)	CR	yes/no
	4.3	Concentration problems (during the self-report)	CR	yes/no
	12	Concentration skills (in general)	CR	9-point scale
	13	Functioning of memory (in general)	CR	9-point scale

Note. MASQ = Mood and anxiety symptoms question; IDS = Inventory of depressive symptoms; WHO-DAS-II = WHO-Disability Assessment Schedule II; 4DSQ = Four-dimensional symptom questionnaire.

[a] Designed by NESDA; not a validated instrument.

Table 2

Descriptive Statistics and Pearson Correlations for Satisficing Indices

Index	Mean	Range	L_{max}	L_{mean}	DRS	ERS	MRS	*G* normed	#missing
L_{max}	4.04 (1.29)	[2, 3]	1.00						
L_{mean}	1.04 (.18)	[1.1, 3.0]	.66	1.00					
DRS	4.36 (3.58)	[0, 23]	.19	.14	1.00				
ERS	.16 (.13)	[0, .80]	−.19	−.31	.05	1.00			
MRS	.25 (.11)	[0, .82]	.10	.18	−.03	−.38	1.00		
G normed	.14 (.08)	[.02, .70]	−.07	−.15	.19	.85	−.38	1.00	
#missing	.18 (1.05)	[0, 27]	.02	.01	.00	.10	−.04	.12	1.00

Table 3

Average Cognitive Symptom Scores for Subgroups

		n	*M (SD)*
Gender			
	Female	1979	0.01 (1.05)
	Male	1002	0.02 (1.01)
Education			
	Basic	199	0.53 (1.08)
	Intermediate	1736	0.08 (1.03)
	High	1046	−0.20 (0.95)
Nationality			
	Dutch	2730	−0.02 (1.00)
	Non-Dutch	251	0.39 (1.15)
Diagnostic status[a]			
	Healthy	1505	−0.54 (0.72)
	Anxious	522	0.12 (0.84)
	Depressed	354	0.60 (0.90)
	Depressed and anxious	564	1.00 (0.93)

Note. [a] "healthy" indicates without a depression or anxiety disorder. Anxious respondents are diagnoses with one or multiple of the following disorders: social phobia (*n* = 547), panic with or without agoraphobia (*n* = 511), agoraphobia (*n* = 152); generalized anxiety disorder (*n* = 389); depressed respondents are diagnosed with either a major or minor depressive disorder (*n* = 868) or dysthymia (*n* = 275).

Table 4

Varimax Rotated Component Loadings From the Principal Component Analysis (PCA) of Validity Indicators

	Component	
	Erratic responding	Repetitive responding
L_{mean}	−.03	**.88**
L_{max}	−.16	**.87**
DRS	**.35**	**.45**
ERS	**.90**	−.26
G normed	**.95**	−.07
#missing	.23	.07
Variance explained	32%	30%
Cronbach's α	.58	.53

Note. Loadings ≥ .35 in bold. Because the oblimin (oblique) rotation method showed a correlation of .02 between the two components, the final PCA solution was obtained using the varimax rotation. MRS was excluded from the PCA because it related negatively to random or inconsistent satisficing. Cronbach's α is derived from the eigenvalue (λ) and the number of variables (M): $\alpha = M\ (\lambda - 1) /\ (M - 1)\lambda$.

Table 5

Mean Satisficing Scores for Different Diagnostic Groups and Corresponding Effect Sizes and Significance Levels for Mean Score Differences

		M (*SD*)		Cohen's *d* A vs. B/C/D	
	n	Erratic	Repetitive	Erratic	Repetitive
A. Healthy	1505	−.12 (.88)	−.12 (.99)	—	—
B. Anxious	522	−.01 (.96)	.17 (1.04)	.12	.29***
C. Depressed	354	.02 (.98)	.01 (.93)	.15	.14
D. Depressed and anxious	564	.33 (1.24)	.15 (1.00)	.44***	.27***

Note. "healthy" indicates without a depression or anxiety disorder; anxious respondents are diagnoses with one or multiple of the following disorders: social phobia (*n* = 547), panic with or without agoraphobia (*n* = 511); agoraphobia (*n* = 152); generalized anxiety disorder (*n* = 389); depressed respondents are diagnosed with either a major or minor depressive disorder (*n* = 868) and/or dysthymia (*n* = 275).

We used Bonferroni adjustment for multiple comparisons. To assess whether the ANOVA and Cohen's *d* were distorted by the skewed distribution of erratic responding, we repeated the analyses using a log transformation of the erratic score (skew = 1.28; kurtosis = 2.15). The results were practically the same.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$ (one tailed)

Table 6

Multiple Regression Analysis Predicting the Two Satisficing Strategies From Cognitive Symptoms and Control Variables

	Erratic responding		Repetitive responding	
Intercept	.71 (.12)	***	.47 (.12)	***
Female gender	−.03 (.04)		.01 (.02)	
Age	.05 (.02)	**	.00 (.02)	
Dutch nationality (vs. non-Dutch)	−.27 (.07)	***	−.21 (.08)	**
Education middle (vs. low)	−.41 (.07)	***	−.21 (.08)	**
Education high (vs. low)	−.53 (.08)	***	−.40 (.08)	***
Cognitive symptoms	.13 (.02)	***	.12 (.02)	***
R^2	.050		.023	
ΔR^2 cognitive symptoms	.016		.013	

Note. Age was standardized.

*$p < .05$, **$p < .01$, ***$p < .001$

Table 7

Regression Coefficients From the Mediation Model Using the Multicategorical Independent Diagnosis Variable, Cognitive Symptom Severity as the Mediating Variable, and One of the Two Satisficing Strategies as the Dependent Variable

Independent variable coding		Erratic responding		Repetitive responding	
Comparison group (baseline)	Group of interest	Total	Indirect	Total	Indirect
Healthy	Anxious	*ns*	n/a	0.27 (.05)	<u>0.07 (0.02)</u>
	Depressed	*ns*	n/a	*ns*	n/a
	Depressed and anxious	0.38 (0.05)	0.12 (0.04)	0.20 (0.05)	<u>0.16 (0.04)</u>
Anxious	Depressed	*ns*	n/a	*ns*	n/a
	Depressed and anxious	0.29 (0.06)	<u>0.07 (0.02)</u>	*ns*	n/a
Depressed	Depressed and anxious	0.26 (0.07)	0.03 (0.01)	*ns*	n/a

Note. “Indirect” is the mediating effect of the specific diagnostic group (vs. comparison group) on the response strategy via cognitive symptom severity. “Total” is the total effect of the specific diagnostic group (vs. comparison group) on the response strategy after controlling for the demographic variables. All coefficients listed in the table are significant at $\alpha = .01$. When total effects are non-significant (*ns*) based on $\alpha = .01$, mediating effects are not applicable (n/a). Indirect (and total) effects that are underlined, were also significant when we re-estimated the model using the clinician-perceived cognitive score.

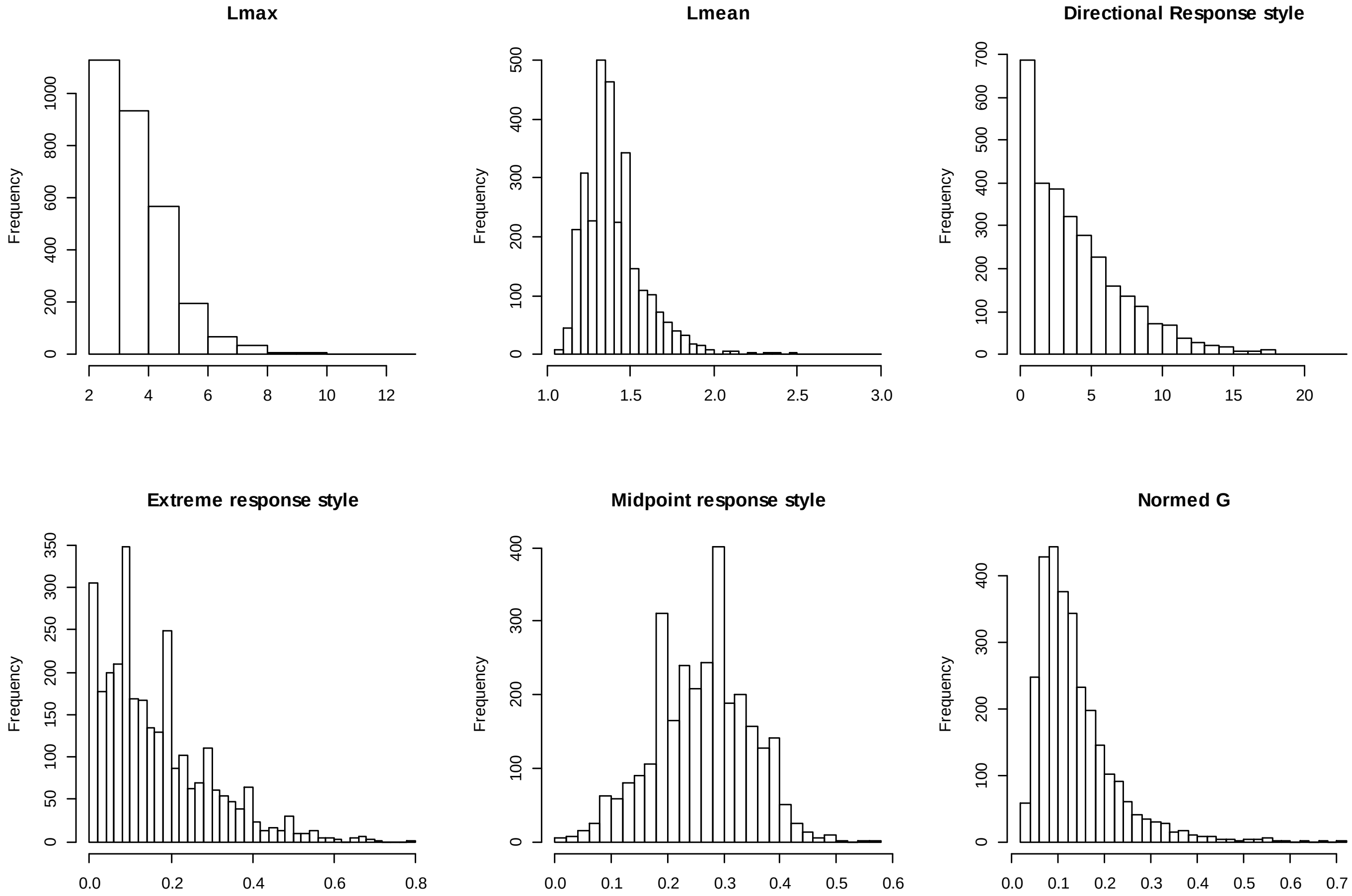

Figure 1. Distributions of response-pattern based validity indicators

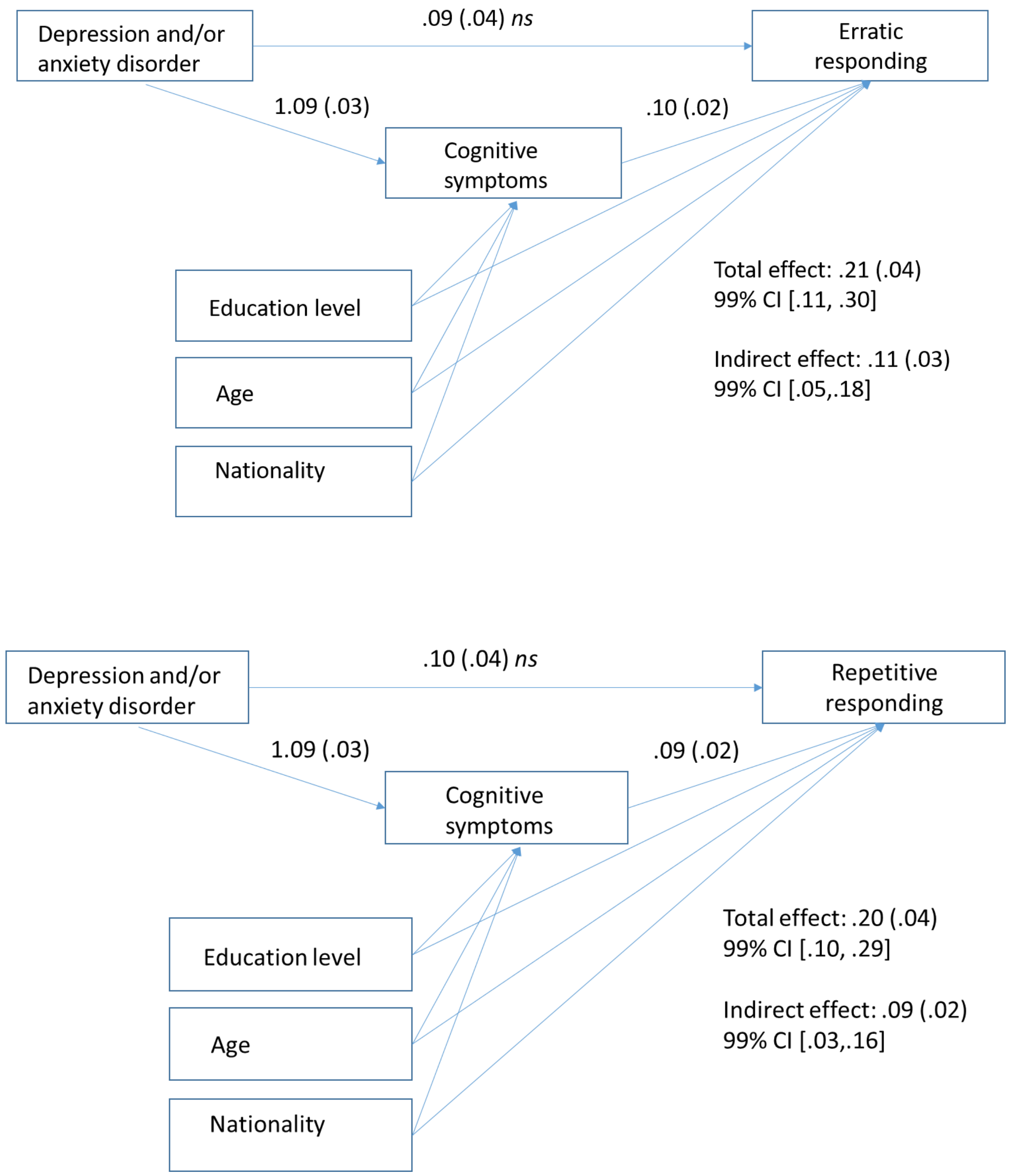

Figure 2. Models representing the mediating effect of cognitive symptoms on erratic responding (upper figure) and repetitive responding (lower figure). “Total effect” is the effect of having a disorder after controlling for the demographic variables

Appendix

Table A1. Component loadings and factor score correlations for the 1-dimensional CATPCA model and 2-dimensional CATPCA model of cogntive symptoms

Scale	Mode	Item content	1-dim. model	2-dim. model	
			Combined	Self	Clinician
IDS (Rush et al., 1996)	SR	Concentration and decision making	.76	**.99**	.02
MASQ-30 (Watson et al., 1995)	SR	Difficulty in taking decisions	.66	**1.02**	−.06
4DSQ distress subscale (Terluin et al., 2006)	SR	Difficulty in thinking clearly	.66	**.98**	.05
WHO-DAS-II (Chwastiak & Von Korff, 2003)	SR	Communication and understanding	.71	**1.01**	−.03
WHODAS-II interview (Buist-Bouwman et al., 2008)	CR	Difficulties in concentrating, memory, and understanding things clearly	.78	**.99**	.02
Evaluation questionnaire for the research assistant[1]	CR	Concentration problems during the interview	.56	.01	**1.00**
		Concentration problems during the self-report	.41	−.03	**1.01**
		Concentration skills (in general)	.73	.04	**.99**
		Functioning of memory (in general)	.60	−.02	**1.01**
			Correlations		
			1		
			.89	1	
			.68	.30	1

Note. Rotation Method: Oblimin with Kaiser Normalization. CATPCA = principal component analysis for categorical data; MASQ = Mood and Anxiety Symptoms Questionnaire; IDS = Inventory of Depressive Symptomatology; WHO-DAS-II = WHO-Disability Assessment Schedule II; 4DSQ = Four-dimensional Symptom Questionnaire. [1]Designed by NESDA; not a validated instrument.