Satisficing in mental health care patients
The effect of cognitive symptoms on self-report data quality
Conijn, J.M.; van der Ark, L.A.; Spinhoven, P.

DOI
10.1177/1073191117714557

Publication date
2020

Document Version
Final published version

Published in
Assessment

License
CC BY-NC

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Satisficing in Mental Health Care Patients: The Effect of Cognitive Symptoms on Self-Report Data Quality

Judith M. Conijn1,2, L. Andries van der Ark1, and Philip Spinhoven2,3

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 health care assessment contexts.

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

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 (1991) 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 noncontent-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, 2000; 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 health care 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 health care 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, 2018; LePagea, Mogge, & Sharpe, 2001; 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 (e.g., Reise & Waller, 1993; Woods, Oltmanns, & Turkheimer, 2008) and clinical samples (Conijn, Emons, et al., 2018; Conijn et al., 2015; 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 (Meade & 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 health care 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; Siebert 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 health care 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 (1991) 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 with 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, & Previle, 2010; Salthouse, 2012). However, not all anxiety disorders may involve cognitive impairment (Castaneda, Tuilio-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 with nonanxious individuals (Bierman, Comijs, Rijmen, Jonker, & Beekman, 2008; Dotson et al., 2014).

This Study

We used Krosnick’s (1991) satisficing theory to identify and explain nonoptimal response strategies in mental health care 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. Our hypotheses were as follows:

Hypothesis 1: Satisficing on the NEO-FFI is more common in respondents with a depression and/or anxiety disorder compared with respondents without these disorders.
Hypothesis 2: Satisficing on the NEO-FFI is positively related to cognitive symptoms, such as problems in concentration, memory, and comprehension.

Hypothesis 3: The severity of cognitive symptoms mediates the positive effect of having a depression and/or anxiety disorder on satisficing.

**Method**

**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 (pp. 21-23). Participants were paid €15 for their participation and compensated for travel costs.

**Measures**

**Depression and Anxiety Disorders.** The lifetime version of the Composite Interview Diagnostic Instrument (Robins et al., 1988) was used to diagnose depressive and anxiety disorders according to the *Diagnostic and Statistical Manual of Mental Disorders–Fourth edition*. The Composite Interview Diagnostic Instrument 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.

---

**Table 1. Items and Scales Used to Assess Cognitive Symptoms of Psychopathology.**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item no./subscale</th>
<th>Measuring Mode</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDS; Rush, Gullion, Basco, Jarrett, and Trivedi (1996)</td>
<td>13</td>
<td>Concentration and decision making</td>
<td>4-Point scale</td>
</tr>
<tr>
<td>MASQ-30; Watson et al. (1995)</td>
<td>25</td>
<td>Difficulty in taking decisions</td>
<td>5-Point Likert-type</td>
</tr>
<tr>
<td>4DSQ distress subscale; Terluin et al. (2006)</td>
<td>12</td>
<td>Difficulty in thinking clearly</td>
<td>5-Point Likert-type</td>
</tr>
<tr>
<td>WHO-DAS-II; Chwastiak and Von Korff (2003)</td>
<td>Subscale (6 items)</td>
<td>Communication and understanding</td>
<td>Sum score</td>
</tr>
<tr>
<td>WHO-DAS-II interview (past month symptoms)</td>
<td>5</td>
<td>Difficulties in concentrating, memory, and understanding things clearly</td>
<td>Yes/no</td>
</tr>
<tr>
<td>Evaluation questionnaire for the research assistant</td>
<td>2.3</td>
<td>Concentration problems (during the interview)</td>
<td>Yes/no</td>
</tr>
<tr>
<td></td>
<td>4.3</td>
<td>Concentration problems (during the self-report)</td>
<td>Yes/no</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Concentration skills (in general)</td>
<td>9-Point scale</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>Functioning of memory (in general)</td>
<td>9-Point scale</td>
</tr>
</tbody>
</table>

Note. MASQ = Mood and Anxiety Symptoms Questionnaire; IDS = Inventory of Depressive Symptoms; WHO-DAS-II = WHO-Disability Assessment Schedule–II; 4DSQ = Four-Dimensional Symptom Questionnaire; SR = self-report; CR = clinician report.

*Designed by the Netherland’s Study of Anxiety and Depression; not a validated instrument.
to careless responding than (Meade & Craig, 2012). We used original item scores of identical answers (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 5-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, 2016). In the next subsections, we describe all six response-pattern-based satisficing indicators. Apart from extreme response style (ERS) and directional response style (DRS), the corresponding 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 ($L_{\text{max}}$) and the average length of the strings of identical answers ($L_{\text{mean}}$). We used original item scores before recoding and ignored missing values. Furthermore, we used both of these indices to assess nondifferentiation. Researchers have found $L_{\text{max}}$ to be somewhat more sensitive to careless responding than $L_{\text{mean}}$ (Meade & Craig, 2012) and $L_{\text{max}}$ may assess severe satisficing. However, $L_{\text{mean}}$ uses all available data and may assess weaker forms of nondifferentiation compared with $L_{\text{max}}$.

Extreme response style. To quantify ERS, we used the percentage of observed 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-type) 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 observed item scores in the middle categories.

Directional response style. Instead of the agreement response style described in Krosnick (1991), 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). 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, Meijer, & Niessen, 2016), 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 satisfiers and nonsatisfiers. This approach reflects that response behavior may range from using optimal strategies to using weak and strong satisficing strategies (Krosnick, 1991). 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 with 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 (Castaneda 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 (Hypothesis 1), we compared the mean satisficing scores across the four diagnostic status categories using multiple $t$ tests. Additional to addressing Hypothesis 1, we also compared the satisficing scores of the three patient groups with each other. We used Bonferroni's correction for multiple comparisons (12 comparisons in total, 6 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 (Hypothesis 2), 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 (Hypothesis 3), 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 four-categorical independent variable. In a first type of model, we used indicator coding for diagnostic status, with the healthy group as the baseline category. This model 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 $\alpha$ 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) with the mediating effect of the disorder on satisficing.

**Additional Analyses.** We conducted sensitivity analyses to assess the extent to which our conclusions regarding hypotheses 2
and 3 would be altered by using either the clinician-perceived cognitive symptom score or the self-reported cognitive symptom score derived from the two-dimensional CATPCA solution (instead of the combined self- and clinician-rated score from the unidimensional solution; see the appendix). Furthermore, to explore which type of persons tend to use specific satisficing strategies, we correlated the satisficing PCA component scores with the NEO-FFI personality traits.

**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 2 shows descriptive statistics and Pearson correlations for the satisficing indicators. As theoretically expected, MRS correlated negatively with ERS, and \( L_{\text{max}} \) and \( L_{\text{mean}} \) were correlated positively. \( G \) normed was highly positively correlated with ERS but negatively correlated with \( L_{\text{mean}} \) 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 repetitive satisficing.

**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 pro-max 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

![Figure 1. Distributions of response-pattern-based validity indices.](image_url)
loadings of $L_{\text{max}}$ and $L_{\text{mean}}$ 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 middle category often was a good response strategy. The correlation pattern between MRS, ERS, and $G_{\text{normed}}$ (see Table 2) 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 3 shows the rotated component loadings. The main difference compared with the previous 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 ($M = 0.0; SD = 1.0; \text{skewness} = 1.66; \text{kurtosis} = 3.95$), whereas the repetitive component score was approximately normally distributed ($M = 0.0; SD = 1.0; \text{skewness} = 1.27; \text{kurtosis} = 3.60$).

### Cognitive Symptoms

Table 4 shows descriptive statistics for the cognitive symptom score for different subgroups. The cognitive symptom score was unrelated to gender, negatively related to education level ($\eta^2 = .03$), and positively related to age ($r = .04$). Respondents with a non-Dutch nationality had a higher mean cognitive symptom score ($d = 0.38$) than Dutch respondents. Compared with 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 with the healthy group (M = −0.54). An analysis of variance showed no substantial effects of specific disorders on the cognitive score (η²s < 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 with healthy respondents. Anxious patients had substantially higher scores for repetitive responding compared with 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 with the depressed group (d = 0.28) and compared with 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 erratic responding and .14 with repetitive responding. 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 2). We included age, nationality, and education level as covariates. After controlling for the covariates, the total effect of the disorder dummy 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 2).

Second, we assessed disorder-specific mediation effects of the cognitive symptom score on satisficing, using diagnostic status as the independent variable (Table 7, column 1). We first compared specific disorder groups with the
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.

<table>
<thead>
<tr>
<th>Independent variable coding</th>
<th>Erratic responding</th>
<th>Repetitive responding</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group of interest</strong></td>
<td>Total</td>
<td>Indirect</td>
</tr>
<tr>
<td>Healthy</td>
<td>Anxious</td>
<td>ns</td>
</tr>
<tr>
<td>Healthy</td>
<td>Depressed</td>
<td>ns</td>
</tr>
<tr>
<td>Healthy</td>
<td>Depressed and anxious</td>
<td>0.38 (0.05)</td>
</tr>
<tr>
<td>Anxious</td>
<td>Depressed</td>
<td>ns</td>
</tr>
<tr>
<td>Anxious</td>
<td>Depressed and anxious</td>
<td>0.29 (0.06)</td>
</tr>
<tr>
<td>Depressed</td>
<td>Depressed and anxious</td>
<td>0.26 (0.07)</td>
</tr>
</tbody>
</table>

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 α = .01. When total effects are nonsignificant (ns) based on α = .01, mediating effects are not applicable (n/a). Indirect (and total) effects that are in italics bold were also significant when we reestimated the model using the clinician-perceived cognitive score and the self-reported cognitive score.

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 satisfying scores between patient groups could be attributed to differences in cognitive symptom severity only to a very small extent.

**Sensitivity Analyses**

We repeated the main analyses using the component scores from the two-dimensional (see the appendix) rather than the one-dimensional CATPCA solution of cognitive symptoms. The two scores separately represented the clinician-perceived and self-reported cognitive symptoms. In the multiple linear regression analyses, we replicated the significant positive effects of cognitive symptoms on satisficing using both measures of cognitive symptoms. The estimated effect of the self-reported cognitive score on erratic responding ($b = .15, p < .01$) was larger than the corresponding effect of the clinician-perceived score ($b = .07, p < .01$). For repetitive responding, the effects of the self-reported cognitive score ($b = .12, p < .001$) and the clinician-perceived cognitive score ($b = .11, p < .001$) were similar.

We then reestimated the general mediation models (Figure 2). Using self-reported cognitive symptoms, we replicated the significant mediating effect for both erratic responding ($b = .10, \text{standard error}[SE] = .04$) and repetitive responding ($b = .07, SE = .03$). Using clinician-rated cognitive symptoms, we replicated the mediating effect for repetitive responding ($b = .03, SE = .01$) but not for erratic responding ($b = .02, SE = .01$). In the disorder-specific mediation models, the results were similar when using either self-reported or clinician-rated cognitive symptoms: We replicated the mediating effects of cognitive symptoms for repetitive responding but not for erratic responding. Table 7 shows the replicated mediating effects in italics.

**Satisficing and Personality**

To assess which personality traits are associated with the use of repetitive and erratic response strategies, we correlated the satisficing scores with the NEO-FFI personality traits. The erratic-responding component had near-zero correlations with each of the traits ($r < |.09|$). The repetitive-responding component correlated weakly with neuroticism ($r = .19$), extraversion ($r = -.11$), openness ($r = -.19$), and agreeableness ($r = -.26$).

**Results for Midpoint Response Style**

Because we excluded the MRS satisficing indicator from the PCA, we repeated the main analyses using MRS as the
dependent variable. Analysis of variance results showed that there were no significant differences in MRS between the diagnostic categories (Hypothesis 1). Multiple regression analysis showed that the cognitive symptom score was significantly related to MRS after accounting for the control variables (Hypothesis 2), but the effect was very small \((b = .01, p < .01)\). We did not conduct a mediation analysis (Hypothesis 3) 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, 2011) 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 health care 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; Meade & Craig, 2012; Niessen et al., 2016). Similar to Meade and Craig (2012), we found two dominant types of satisficing strategies: erratic (i.e., extreme or inconsistent) responding and repetitive responding.

**Main Results**

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 with healthy respondents. The effect size and type of satisficing strategy used differed across diagnostic categories. Anxious respondents used more repetitive responding compared with 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% to 2%) but larger than the variance explained by demographic characteristics, such as education level. The low percentages of explained variance in satisficing may partly be due to low reliability of the satisficing scores.

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 replicated using both the clinician-rated and the self-reported 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 or the self-reported 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—indeed, 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 decisional anhedonia explains additional variance in satisficing and whether it could provide an explanation for the low variance explained in satisficing scores in our current study.

Our results suggest that nonoptimal response strategies may be common in mental health care 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. The NESDA study includes volunteers and a substantial subgroup with no current mental disorder. In other mental health care 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 questionnaire could be compared between different assessment settings.

**Limitations and Future Research**

An important question relating to the validity of our study’s conclusions is the degree to which our measure of cognitive symptom severity was contaminated by satisficing. Although cognitive symptoms were measured by both self-report and clinician ratings, the combined score appeared to mainly reflect self-reported problems. Our sensitivity analyses also showed some evidence for a bias in the assessment of cognitive symptoms: When we used the clinician-perceived cognitive score instead of the combined patient-clinician score, effect sizes were smaller, and the mediating effects of cognitive symptom severity could only partly be replicated. So, possibly, the cognitive score was affected by satisficing or other response biases, such as malingering, and the regression effects in our main results may be biased.

However, this is only one possible explanation for the inconsistent results. Other plausible explanations are related to the quality of the clinician rating: (a) the clinical research assistants had to indirectly infer cognitive problems from a respondent’s functioning during the interview; (b) research assistants could not compare the cognitive skills of patients with respect to their previous (nondepressed) functioning, so cognitive symptoms may not only reflect problems related to psychopathology; and (c) the rating instrument was not validated and reliability was low (α = .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, 2011). Replication research that uses a high-quality objective measure of cognitive functioning is needed to estimate effect sizes correctly.

This study has several other limitations. First, 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% to 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 health care research and practice. A second 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 (38% 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 ERS. 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, 2017; Wanders, Wardenaar, Penninx, Meijer, & de Jonge, 2015).

Finding out which individual characteristics cause a person to use a specific satisficing strategy remains a topic for future research. The importance of the topic goes beyond health care research. Our explorative results showed that repetitive responding was positively related to disagreeableness, tentatively suggesting that repetitive responding may result from uncooperative behavior. However, the correlations between repetitive responding and disagreeableness might be biased by satisficing. Erratic responding was not substantially related to any of the personality traits. Future research pursuing a behavioral analysis of satisficing may use the following analytic strategies. First, one may use a mixed-effects explanatory IRT model (de Boeck & Wilson, 2004) that treats each item score as a consecutive satisficing indicator. This model allows for the inclusion of explanatory variables to study which between-person variables (e.g., personality traits, intelligence) and which within-person variables (e.g., item difficulty, previous response) induce satisficing. Another idea for future research is to adapt decision-making models from the behavioral literature for quantifying 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 health care research merits further attention. Future research ought to address the following questions: (a) To what extent do different health care assessment contexts induce satisficing strategies? (b) At what level do cognitive problems necessitate the use of rating scales instead of self-report measures? (c) To what extent do satisficing strategies bias test scores and affect research conclusions?
## Appendix

Component Loadings and Factor Score Correlations for the One-Dimensional CATPCA Model and Two-Dimensional CATPCA Model of Cognitive Symptoms.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mode</th>
<th>Item content</th>
<th>One-dimensional model</th>
<th>Two-dimensional model</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDS; Rush et al. (1996)</td>
<td>SR</td>
<td>Concentration and decision making</td>
<td>0.76</td>
<td>0.99</td>
</tr>
<tr>
<td>MASQ-30; Watson et al. (1995)</td>
<td>SR</td>
<td>Difficulty in taking decisions</td>
<td>0.66</td>
<td>1.02</td>
</tr>
<tr>
<td>4DSQ distress subscale; Terluin et al. (2006)</td>
<td>SR</td>
<td>Difficulty in thinking clearly</td>
<td>0.66</td>
<td>0.98</td>
</tr>
<tr>
<td>WHO-DAS-II; Chwastiak and Von Korff (2003)</td>
<td>SR</td>
<td>Communication and understanding</td>
<td>0.71</td>
<td>1.01</td>
</tr>
<tr>
<td>WHO-DAS-II interview; Buist-Bouwman et al. (2008)</td>
<td>CR</td>
<td>Difficulties in concentrating, memory, and understanding things clearly</td>
<td>0.78</td>
<td>0.99</td>
</tr>
<tr>
<td>Evaluation questionnaire for the research assistant*</td>
<td>CR</td>
<td>Concentration problems during the interview</td>
<td>0.56</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Concentration problems during the self-report</td>
<td>0.41</td>
<td>−0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Concentration skills (in general)</td>
<td>0.73</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Functioning of memory (in general)</td>
<td>0.60</td>
<td>−0.02</td>
</tr>
</tbody>
</table>

Correlations

\[
\begin{array}{ccc}
1 & .89 & 1 \\
.89 & 1 & .30 \\
.68 & .30 & 1
\end{array}
\]

Note. CATPCA = categorical principal components analysis; MASQ = Mood and Anxiety Symptoms Questionnaire; IDS = Inventory of Depressive Symptomatology; WHO-DAS-II = WHO-Disability Assessment Schedule–II; 4DSQ = Four-Dimensional Symptom Questionnaire; SR = self-report; CR = clinician report. Rotation Method: Oblimin with Kaiser normalization.

*Designed by the Netherland’s Study of Anxiety and Depression; not a validated instrument.
Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: 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).

Notes
1. Several alternative indices can be used to assess random responding, such as person-fit statistic $I_z$ or the Mahalanobis distance (e.g., Meade & Craig, 2012; Niessen et al., 2016). 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. (2016), we choose the $G$ person-fit statistic: (a) it imposes a less restrictive model on the data than the $I_z$ index and (b) Niessen et al. (2016) found that $G$ performed equally well compared with $I_z$ statistic but better than the Mahalanobis distance.

2. Alternatively, satisfying may be a categorical construct, as suggested in research investigating careless responding (Kam & Meyer, 2015; Meade & Craig, 2012). Following these studies, in preliminary analyses, we used latent class profile analysis to assess whether we could identify latent satisfying groups based on the seven validity indicators. Results showed that model fit consistently improved (up to nine 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 satisfying would be more appropriate.

References
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:...


