Too little, just right or too much? Assessing how people evaluate their conscientiousness levels
Dupre, Sofie; Heyde, Fien; Vergauwe, Jasmine; Wille, Bart

Published in:
Personality and Individual Differences

DOI:
10.1016/j.paid.2022.111789

Publication date:
2022

License:
CC BY-NC-ND

Document Version:
Accepted author manuscript

Citation for published version (APA):
Too little, just right or too much? Assessing how people evaluate their conscientiousness levels

Article in Personality and Individual Differences - October 2022
DOI: 10.1016/j.paid.2022.111789

CITATION
1

READS
123

4 authors:

Sofie Dupré
Ghent University
1 PUBLICATION 1 CITATION
SEE PROFILE

Fien Heyde
Ghent University
6 PUBLICATIONS 5 CITATIONS
SEE PROFILE

Jasmine Vergauwe
Vrije Universiteit Brussel
33 PUBLICATIONS 467 CITATIONS
SEE PROFILE

Bart Wille
Ghent University
81 PUBLICATIONS 1,911 CITATIONS
SEE PROFILE

All content following this page was uploaded by Sofie Dupré on 11 January 2023.
The user has requested enhancement of the downloaded file.
Too little, just right or too much? Assessing how people evaluate their conscientiousness levels

Sofie Dupré, Fien Heyde, Jasmine Vergauwe, & Bart Wille

Ghent University

Abstract

Recent advances in personality theory and research have led to the introduction of the “Too-much-of-a-good-thing-effect” in the relationship between conscientiousness and desirable outcomes, challenging the “more is better” idea that has been dominating research on this trait for a long time. Thus, the question arises as to how people evaluate their conscientiousness levels themselves, more specifically, whether they regard their trait levels as “too little”, “the right amount”, or “too much”. The current study describes how an existing personality inventory can be adjusted to explore such evaluations of conscientiousness levels by incorporating a too little/too much response format. The structural characteristics of this new assessment approach are examined and compared against responses that are collected using a traditional Likert rating format asking people to describe themselves. Results show that – in this sample ($N = 367$) – about 11% of participants evaluated their conscientiousness as adequate, whereas the majority (75%) indicated it to be too high. Further, the “right amount” of conscientiousness was most frequently associated with a 7 on a 9-point Likert scale, while very high Likert-scale ratings of 9 were regarded as “too much” in over three-fourth of the ratings. Implications and directions for future research are discussed.

Keywords: personality assessment, conscientiousness, too-much-of-a-good-thing, too little/too much (TLTM) scale
1. Introduction

High conscientiousness is a reliable correlate of multiple desirable outcomes in different areas of people’s lives, such as physical and psychological well-being (Friedman et al., 2010), academic success (Mammadov, 2021), and work performance (Wilmot & Ones, 2019). Although these conscientiousness-outcome relations have been deemed positive and linear in most research (i.e., “more is better”), contradictory evidence from recent lines of theoretical and empirical work indicates that there may be such a thing as being too conscientious (e.g., Carter et al., 2014, 2016; Denissen et al., 2018; Le et al., 2011). This phenomenon – also referred to as the “Too-much-of-a-good-thing-effect” (TMGT effect; Pierce & Aguinis, 2013) – raises the question how people feel about or evaluate their conscientiousness levels. While current personality assessment instruments provide valuable insights into people’s positions on the underlying trait dimension, they cannot capture such evaluations (i.e., whether people regard their trait levels as “too little”, “the right amount”, or “too much”). The current study is a first effort to explore how people evaluate their conscientiousness levels by introducing the too little/too much rating scale (Kaiser & Kaplan, 2005) in personality assessment.

1.1 Conscientiousness from a TMGT Perspective

Although the TMGT-effect manifests in different life domains (Pierce & Aguinis, 2013), evidence for this effect regarding conscientiousness has mainly been reported in the work setting, where curvilinear relationships have been demonstrated with various indicators of job performance. That is, several studies demonstrate that employees with moderate conscientiousness levels tend to perform better than those with very high levels (Carter et al., 2014; Denissen et al., 2018; Le et al., 2011). Highly conscientious employees may be so perfectionist that they waste time on unimportant details. Importantly, Le et al. (2011) found that higher levels of conscientiousness only become detrimental to performance in low-
complexity jobs, which demonstrates the importance of the work context in explaining the nature of the curvilinear relationship. Similarly, Denissen et al. (2018) showed that high conscientiousness in employees has a detrimental effect on performance when conscientiousness exceeds the level their job demands, which provides further evidence for the significance of a good match between personality and the job context.

While the issue of curvilinearity has been mainly addressed in research on the conscientiousness-performance relationship, it is not exclusive to the work context. For example, Carter et al. (2016) found a similar trend in the relation between conscientiousness and psychological well-being. Taken together, these findings raise the possibility that people who report higher conscientiousness levels potentially experience these trait levels as a hindrance in life, and may feel that lower levels are more desirable.

1.2 Assessing People’s Evaluations of Conscientiousness

For decades, psychologists have been in pursuit of valid and reliable ways to assess people’s personality, but despite the emergence of a multitude of instruments, none have aimed to capture the way people evaluate their reported trait levels. While Likert scales are designed to grasp the person’s actual position on the underlying trait, they are not intended to reveal discrepancies between actual trait levels and desired trait levels. For example, people who rate themselves as highly conscientious on a Likert scale may regard these levels as either adequate or potentially “too high”. Similarly, people with lower Likert-scale rated levels are not necessarily dissatisfied with these levels. Previous efforts have been made toward better operationalizing potential adaptive and maladaptive functioning at both poles of a trait dimension, such as the Five Factor Form (FFF) and the Sliderbar Inventory (SI) (Rojas & Widiger, 2018). However, like the Likert scale, these scales are still aimed at – more accurately – describing people’s trait levels.
Similar remarks were made previously by Kaiser and Kaplan (2005), who aimed to assess deficiency and/or excess in leadership behavior. In their study, they suggested a new rating scale format: the too little/too much rating scale, ranging from -4 (too little) to 0 (the right amount) and +4 (too much) (Kaiser and Kaplan, 2005). The usefulness of this scale lies in its ability to tap into perceived overdoing or underdoing, which are areas of functioning not captured by the Likert scale rating format. In the current study, our goal is to introduce this scale in personality assessment to explore employees’ evaluations of their conscientiousness levels.

The current study presents the development and validation of a personality instrument that focuses on mapping people’s evaluations of their conscientiousness levels (“too much”, “too little”, or “the right amount”) rather than the actual conscientiousness levels (1 = very low; 9 = very high). The characteristics of this instrument will be investigated in two ways. First, the factor structure of the TLTM-rated version will be compared to that of the traditional Likert-rated version of the instrument. Second, associations between ratings on TLTM- and Likert-versions of the instrument will be investigated to build an understanding of how people evaluate different conscientiousness trait levels.

2. Method

2.1 Participants and Procedure

Data were collected using an online cross-sectional survey. Participants were recruited through the personal networks of the researchers and comprise employed people coming from a variety of businesses and organizations (e.g., schools, elderly care companies). The study comprised 367 Belgian adults (72% female) with an average age of 39.68 years (SD = 11.87). The majority of participants held either a high school degree (43%) or a bachelor’s degree (43%).
2.2 Instrument Development

Participants were asked to rate their level of conscientiousness twice: first using a 9-point Likert scale ranging from 1 (totally disagree) to 9 (totally agree), and second using a 9-point TLTM-scale ranging from -4 (much too little) to +4 (much too much). For the TLTM scale labels, we decided to stay as close as possible to the original labels by Kaiser & Kaplan (2005) which are: -4 (much too little), -1 (barely too little), 0 (the right amount), +1 (barely too much), and +4 (much too much). However, since Dutch translations for the -1 and +1 labels do not make much sense linguistically, we labeled -2 (too little) and +2 (too much) instead. For the Likert scale, we chose a 9-point scale to ensure equal scale coarseness (Vergauwe et al., 2017) and used a similar pattern of labels: 1 (totally disagree), 3 (disagree), 5 (neutral), 7 (agree), and 9 (totally agree). The Likert scale was presented first for two reasons. First, when attempting to grasp people’s evaluations, the approach that makes the most sense for participants is to let them indicate their actual trait levels first (descriptive), and subsequently allow them to reflect on their evaluations of these trait levels. Similar approaches have been used previously in research on personality change goals, in which participants are presented with a traditional personality inventory first, and are then asked to reflect on their desires to change certain traits using a modified version of the personality inventory (e.g., Hudson & Roberts, 2014). A second reason for this order is that, in line with Vergauwe et al. (2017), the evaluative component of the TLTM format was expected to have a stronger influence on the Likert scale ratings than vice versa.

For both assessments, a Dutch translation of the NEO-PI-R itemset (Costa & McCrae, 1992; Hoekstra et al., 2007), including 48 items to measure six conscientiousness facets (i.e., C1: Order, C2: Competence, C3: Dutifulness, C4: Achievement, C5: Self-Discipline, and C6: Deliberation) was used as a starting point. Given that the NEO-PI-R items are designed to be rated using a Likert scale, a first step consisted of checking all 48 items in terms of their “fit”
with a TLTM-type rating scale. Specifically, as also highlighted by similar work in leadership assessment, items needed to be evaluated against two criteria (Vergauwe et al., 2017). First, a TLTM-rated item needs to be positively keyed, meaning that it describes the higher range of conscientiousness and does not need reverse scoring. Second, since the TLTM scale implies an evaluative response, it is important that the item itself remains purely descriptive, avoiding any evaluative content (Vergauwe et al., 2017). For example, the item “I feel well prepared to deal with life” was changed to “I feel prepared to deal with life”. Items that could not be altered accordingly were removed, which resulted in a final selection of 24 out of the original 48 items, with each facet being covered by four items. Example items for each facet include “I am neat and tidy” (Order), “I feel prepared to deal with life” (Competence), “I adhere to my ethical principles” (Dutifulness), “I work hard to achieve my goals” (Achievement Striving), “I have the ability to motivate myself to get the job done” (Self-Discipline), and “I am cautious and deliberate” (Deliberation). Table 1 gives an overview of descriptive statistics, correlations, and internal consistencies. Chronbach alphas for each Likert subscale were good, ranging between \( \alpha = .73 \) (Dutifulness) and \( \alpha = .88 \) (Self-Discipline). For the TLTM scale, the internal consistency for the facet Dutifulness was slightly lower (\( \alpha = .63 \)), while for the other facets it ranged between \( \alpha = .78 \) (Competence) and \( \alpha = .85 \) (Order). There were positive correlations (average \( r = .52 \)) between all corresponding facets of both scales, indicating that higher Likert conscientiousness ratings are generally associated with TLTM ratings toward the ‘too much’ end of the scale. The lowest correlation was found between Likert and TLTM ratings of Dutifulness (\( r = .27 \)). A potential explanation is that Likert scores for this facet were exceptionally high in comparison to other facets (i.e. they ranged between 4.75 and 9), which may have may have attenuated the correlation coefficient.

To get a sense of the response patterns, we categorized TLTM ratings as the right amount when conscientiousness scores were within +/-3 standard errors of measurement
around 0, since these scores are statistically indistinguishable from 0 at $p < .001$ (Ghiselli et al., 1981). Conscientiousness scores exceeding (below) this range were categorized as too much (too little). As shown in Table 2, ratings on the TLTM scale were most frequent in the “too much” range for all conscientiousness facets, indicating that the majority of participants felt that their current conscientiousness levels were higher than what was considered ideal.

3. Results

3.1 Factor Structures

In further evaluating the characteristics of the TLTM-rated personality instrument, we first examined its factor structure and compared it to the structure of the traditional Likert-type rating scale. Before turning to the TLTM-version, we re-examined the Likert version of the instrument given that (a) it includes only a selection of the original NEO-PI-R itemset, (b) the Likert scale was slightly adjusted (i.e. nine instead of the typical five response options), and (c) in some cases the items were slightly adapted to make them also appropriate for the TLTM-scale.

3.1.1 Likert scale.

To explore the expected multidimensional structure at the facet level we conducted Exploratory Structural Equation Modeling (ESEM; Asparouhov & Muthén, 2009) using Mplus version 8.4 (Muthén & Muthén, 2017). This method has been suggested to be more appropriate for personality items than confirmatory factor analysis (CFA), because it is more flexible in allowing cross-loadings in contrast to the rather unrealistic (and theoretically undesirable) assumption of simple structures underlying CFA models (Marsh et al., 2010). We chose MLR estimation because the data were not normally distributed. Model fit was first evaluated based on RMSEA, CFI and TLI fit indices. Although different cut-offs have been described in the literature, an often used rule of thumb is proposed by Hu and Bentler (1999) who asserted that a cutoff value “close to .06” is needed for RMSEA and values “close to
.95” are needed for TLI and CFI to conclude that there is a relatively good fit between the hypothesized model and the observed data. Applying these criteria to the current data indicates relatively good results for RMSEA (i.e., .060) and – to a lesser extent – CFI (i.e., .93), but poor fit for TLI (i.e., .87). However, in addition to these fit indices, model fit is also evaluated on a more theoretical ground by inspecting the pattern of factor loadings and evaluating to what extent these factors indeed capture the concepts that are assumed to be underlying. As can be seen from Table 1 in the Supplementary Material, participants’ responses on the Likert scale showed a clear six-factor structure with items generally loading primarily on their intended factors. Two exceptions were that Dutifulness items 11 and 12 loaded more strongly onto the factor representing Deliberation1.

3.1.2 TLTM scale.

A similar approach was followed for the TLTM-rated items. We ran a ESEM model allowing cross-loadings for the TLTM-rated personality items. An inspection of the fit indices shows acceptable results for RMSEA (i.e., .063) and CFI (i.e., .93), but again, poor fit for TLI (i.e., .88). However, as can be seen from Table 3, the pattern of factor loadings was highly similar compared to the Likert version: overall, participants’ responses were organized in terms of six clearly distinguishable conscientiousness facets. The only exception was that item 11 loaded more strongly on the factor representing Deliberation than on its intended factor representing Dutifulness2.

3.2 Associations between Likert- and TLTM-rated Responses

Next, we proceeded by investigating the associations between Likert- and TLTM-rated responses. We first plotted the relationship between each Likert item rating and its respective TLTM item score for all 24 items. Across all items, participants indicating the

---

1 We additionally ran the more stringent CFA approach which resulted in a model with an overall better fit. Factor loadings and fit estimates are presented in Table 2 in the Supplementary Material.
2 Again, we also ran the CFA approach which resulted in a model with an overall better fit. Factor loadings and fit estimates are shown in Table 3 in the Supplementary Material.
right amount on the TLTM scale reported Likert ratings between 3 and 9 in 37.5% of the items. For the remaining items, the right amount on the TLTM scale corresponded to a Likert scale range between 4 and 9 (20.83%), 5 and 9 (16.66%), 1 and 9 (12.50%), and 2 and 9 (12.50%). Importantly, more narrow ranges (e.g., 5 – 9) were potentially attributable to the small number of participants reporting Likert scores of 4 and below on these items. Across all 24 items, the right amount was most commonly associated with a 7 on the Likert scale (39.85%), indicating that more conscientiousness is not always evaluated as “better”. In fact, the highest Likert score (9) corresponded to “too much” in 76% of the ratings, and was only associated with the right amount in 22.6% of the cases. Due to a lack of Likert responses on the lower end of the scale, we were unable to explore associations between Likert scores of 1 and 2 with the TLTM scale (i.e. only 12 participants indicated a 1 and 2 on the lowest rated Likert item).

The general pattern of associations between responses on both scales is illustrated for item 1 (“I am neat and tidy”) in Figure 1. Panel A shows that for item 1, the right amount (i.e., 0 on the TLTM-scale) was associated with a broad range of Likert scale options (i.e., 3, 4, 5, 6, 7, 8, or 9). Panel B provides a more detailed overview of the frequencies of each response on item 1, showing that the right amount was most frequently associated with a Likert score of 7, and less with an 8 or 9. Similar analyses can be reported for the remaining items (see Supplementary Material).

4. Discussion

In the past decade, theory and research on personality have been experiencing a shift from a “more is better” paradigm to the exploration of potential curvilinear effects, in support of the idea that some people may have “too much” of a desirable trait. Regarding the trait conscientiousness, most curvilinear relations have been found when looking at predictive effects for job performance (Carter et al., 2014; Le et al., 2011), though similar effects are
being discovered in different areas of people’s lives such as psychological well-being (Carter et al., 2016). This raises the question to which degree certain conscientiousness levels are evaluated by respondents as too little, the right amount, or too much. The present study was a first effort to explore these evaluations, by applying the too little/too much rating scale previously used in leadership assessment (Kaiser & Kaplan, 2005; Vergauwe et al., 2017). Doing so, the focus shifts from mapping actual trait levels to mapping people’s evaluations of these levels in terms of underdoing (too little), overdoing (too much), or the right amount. In the current sample, TLTM ratings were most frequent in the “too much” range for all conscientiousness facets, with much less people evaluating their conscientiousness levels as “too little” or “the right amount”.

We started this exploration by departing from a well-established personality inventory, the NEO-PI-R. In a first step, a selection of (slightly adapted) conscientiousness items was made that can be rated using both Likert and TLTM response formats. Given these adaptations, it was important to first evaluate the characteristics of the TLTM-rated personality items by investigating the underlying factor structure. When strictly considering commonly used cut-offs for fit indices, the fit of an ESEM-model was suboptimal for both Likert-rated and TLTM-rated items. However, an inspection of the pattern of factor loadings revealed that – similar to the Likert-rated items – TLTM-rated items were organized in terms of six clearly distinguishable conscientiousness facets. This is important, as one concern could be that the evaluative nature of the TLTM rating scale could introduce a strong evaluative component in the personality assessment, potentially distorting the factor structure.

Next, we proceeded by investigating the relationships between the ratings on both scales. This analysis revealed that, on average, the right amount of conscientiousness corresponded to ratings of 7 on the 9-point Likert scale. Additionally, Likert-scale ratings of 9 were regarded “too much” in over three-fourths of the ratings. Taken together, these results
clearly indicate that more conscientiousness is not always regarded as better, and people can indeed experience these higher trait levels as “too much”.

4.1 Limitations and Future Directions

Several limitations of this study can be mentioned. First, the current sample obtained a rather high average conscientiousness score ($M = 6.98, SD = 0.76$ on the 9-point Likert scale). As a consequence, there was insufficient data at the lower end of this scale to explore potential associations with TLTM ratings. As such, it remains an open question as to whether low Likert ratings can also be evaluated as “the right amount” (or not). Future research could try to replicate our findings in a new sample, ensuring more heterogeneity in participants’ conscientiousness levels. Second, raters were asked to rate the same conscientiousness items twice, using two different rating scales. For multiple theoretical reasons we presented the Likert scale format first. However, future research could use a counterbalanced design to test whether the order may affect the results.

An important question that arises from these findings relates to the frame-of-reference that was used by participants when evaluating their trait levels. While some may have evaluated their levels of conscientiousness in light of their private lives, others may have drawn from work experiences, or perhaps a combination of both. Several studies on the TMGT effect in organizational psychology already highlight the importance of context (Denissen et al., 2018; Le et al., 2011), in that some jobs call for higher trait levels than others. Consequently, it is possible that the ideal level of a trait – as well as people’s evaluation thereof – differs depending on the context that people draw from when rating this type of items. Similar to what has been done in “traditional” (i.e. Likert-type) personality assessment (e.g., Swift & Peterson, 2019), future research can explore how contextualizing TLTM-assessments of personality may help to better understand people’s evaluations of their trait levels relative to specific contextual demands. For example, future studies could
explicitly ask employees to evaluate their trait levels in relation to their specific work
environment, and explore how these evaluations would relate to contextualized Likert-scale
ratings.

Nevertheless, this study is the first to examine the potential of the TLTM-scale to
grasp people’s personal *evaluations* of their personality trait levels. While these initial results
are promising, several paths can be further explored. First, future studies could extend our
approach to other big five traits. Considering the existence of the TMGT effect for other traits
such as emotional stability (Le et al., 2011), and extraversion (Ames & Flynn, 2007; Grant,
2013), it is possible that similar patterns in evaluations of those traits would be observed.
However, the right amount could also be associated with different Likert-scale levels for
different traits. Furthermore, since the TLTM format has been proven to add incremental
validity over Likert ratings in leadership research (Vergauwe et al., 2017), researchers could
explore similar predictive qualities of the TLTM scale in applied personality research. In its
current self-report format, the TLTM scale could be particularly useful in a development or
coaching setting, and future studies could explore how TLTM personality ratings relate to
external criteria such as people’s personality development goals. Importantly, because of
their evaluative nature, self-reported TLTM ratings would be less appropriate to use in a
testing context that is highly sensitive to socially desirable responding (e.g., employee
selection). However, an interesting avenue could be to collect other-ratings of personality –
similar to what has been done in traditional personality assessment (Connelly & Ones, 2010)
– and explore correlates of these ratings. For example, managers’ evaluation of their
employees’ conscientiousness using the TLTM scale could be used to predict employee
performance.

5. Declarations of interest

   Declarations of interest: none
6. Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
6. Literature


Table 1

Descriptive Statistics, Variable Intercorrelations and Internal Consistencies

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conscientiousness TLTM</td>
<td>335</td>
<td>0.64</td>
<td>0.76</td>
<td>.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Order TLTM</td>
<td>335</td>
<td>0.38</td>
<td>0.88</td>
<td>.80***</td>
<td>(.82)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Competence TLTM</td>
<td>335</td>
<td>0.63</td>
<td>1.04</td>
<td>.74***</td>
<td>.53***</td>
<td>(.80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Dutifulness TLTM</td>
<td>335</td>
<td>1.13</td>
<td>0.91</td>
<td>.74***</td>
<td>.49***</td>
<td>.45***</td>
<td>(.73)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Achievement Striving TLTM</td>
<td>335</td>
<td>0.66</td>
<td>1.07</td>
<td>.76***</td>
<td>.58***</td>
<td>.44***</td>
<td>.52***</td>
<td>(.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Self-Discipline TLTM</td>
<td>335</td>
<td>0.71</td>
<td>1.08</td>
<td>.81***</td>
<td>.59***</td>
<td>.47***</td>
<td>.56***</td>
<td>.61***</td>
<td>(.88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Deliberation TLTM</td>
<td>335</td>
<td>0.36</td>
<td>1.08</td>
<td>.63***</td>
<td>.42***</td>
<td>.43***</td>
<td>.36***</td>
<td>.25***</td>
<td>.36***</td>
<td>(.76)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Conscientiousness Likert</td>
<td>357</td>
<td>6.98</td>
<td>0.76</td>
<td>.47***</td>
<td>.31***</td>
<td>.45***</td>
<td>.15**</td>
<td>.38***</td>
<td>.39***</td>
<td>.37***</td>
<td>(.89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Order Likert</td>
<td>357</td>
<td>6.71</td>
<td>1.00</td>
<td>.26***</td>
<td>.39***</td>
<td>.27***</td>
<td>.07</td>
<td>.22***</td>
<td>.16**</td>
<td>.18**</td>
<td>.70***</td>
<td>(.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Competence Likert</td>
<td>357</td>
<td>6.88</td>
<td>1.29</td>
<td>.31***</td>
<td>.13*</td>
<td>.60***</td>
<td>.07</td>
<td>.15**</td>
<td>.22***</td>
<td>.21***</td>
<td>.71***</td>
<td>.43***</td>
<td>(.78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Dutifulness Likert</td>
<td>357</td>
<td>7.59</td>
<td>0.78</td>
<td>.26***</td>
<td>.11*</td>
<td>.19***</td>
<td>.27***</td>
<td>.16**</td>
<td>.17**</td>
<td>.24***</td>
<td>.67***</td>
<td>.35***</td>
<td>.34***</td>
<td>(.63)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Achievement Striving Likert</td>
<td>357</td>
<td>6.88</td>
<td>1.18</td>
<td>.40***</td>
<td>.33***</td>
<td>.26***</td>
<td>.19***</td>
<td>.63***</td>
<td>.32***</td>
<td>.05</td>
<td>.68***</td>
<td>.44***</td>
<td>.30***</td>
<td>.36***</td>
<td>(.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Self-Discipline Likert</td>
<td>357</td>
<td>7.19</td>
<td>1.08</td>
<td>.39***</td>
<td>.23***</td>
<td>.26***</td>
<td>.15**</td>
<td>.37***</td>
<td>.59***</td>
<td>.13*</td>
<td>.76***</td>
<td>.39***</td>
<td>.43***</td>
<td>.45***</td>
<td>.57***</td>
<td>(.84)</td>
<td></td>
</tr>
<tr>
<td>14. Deliberation Likert</td>
<td>357</td>
<td>6.61</td>
<td>1.22</td>
<td>.27***</td>
<td>.10</td>
<td>.22***</td>
<td>.02</td>
<td>.02</td>
<td>.14*</td>
<td>.67***</td>
<td>.65***</td>
<td>.34***</td>
<td>.34***</td>
<td>.46***</td>
<td>.18***</td>
<td>.36***</td>
<td>(.84)</td>
</tr>
</tbody>
</table>

Note: Values between brackets on the diagonal show the Chronbach’s Alpha (α) of the relevant variable. Correlations between corresponding facets of both scales are in bold. TLTM = too little/too much.

*p < .05, **p < .01, ***p < .001.
Table 2

*Frequencies (%) of TLTM Conscientiousness Categories (N=335)*

<table>
<thead>
<tr>
<th>Category</th>
<th>Too Little</th>
<th>The right Amount</th>
<th>Too Much</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conscientiousness</td>
<td>13.4</td>
<td>11.0</td>
<td>75.5</td>
</tr>
<tr>
<td>C1: Order</td>
<td>24.5</td>
<td>21.8</td>
<td>53.7</td>
</tr>
<tr>
<td>C2: Competence</td>
<td>18.5</td>
<td>11.6</td>
<td>69.9</td>
</tr>
<tr>
<td>C3: Dutifulness</td>
<td>5.1</td>
<td>9.9</td>
<td>85.1</td>
</tr>
<tr>
<td>C4: Achievement Striving</td>
<td>20.0</td>
<td>11.6</td>
<td>68.4</td>
</tr>
<tr>
<td>C5: Self-Discipline</td>
<td>17.9</td>
<td>15.8</td>
<td>66.3</td>
</tr>
<tr>
<td>C6: Deliberation</td>
<td>26.3</td>
<td>15.5</td>
<td>58.2</td>
</tr>
</tbody>
</table>

*Note: TLTM ratings were categorized as the right amount when conscientiousness scores were within +/- 3 standard errors of measurement around 0, since these scores are statistically indistinguishable from 0 at p < .001 (Ghiselli et al., 1981). Conscientiousness scores exceeding this range were categorized as too much. Conscientiousness scores below this range were categorized as too little.*
Table 3

**ESEM standardized factor loadings of the TLTM scale items**

<table>
<thead>
<tr>
<th>Items</th>
<th>Factor loading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>Order</strong></td>
<td></td>
</tr>
<tr>
<td>c1 TLTM</td>
<td></td>
</tr>
<tr>
<td>c2 TLTM</td>
<td></td>
</tr>
<tr>
<td>c3 TLTM</td>
<td></td>
</tr>
<tr>
<td>c4 TLTM</td>
<td></td>
</tr>
<tr>
<td><strong>Competence</strong></td>
<td></td>
</tr>
<tr>
<td>c5 TLTM</td>
<td></td>
</tr>
<tr>
<td>c6 TLTM</td>
<td></td>
</tr>
<tr>
<td>c7 TLTM</td>
<td></td>
</tr>
<tr>
<td>c8 TLTM</td>
<td></td>
</tr>
<tr>
<td><strong>Dutifulness</strong></td>
<td></td>
</tr>
<tr>
<td>c9 TLTM</td>
<td></td>
</tr>
<tr>
<td>c10 TLTM</td>
<td></td>
</tr>
<tr>
<td>c11 TLTM</td>
<td></td>
</tr>
<tr>
<td>c12 TLTM</td>
<td></td>
</tr>
<tr>
<td><strong>Achievement Striving</strong></td>
<td></td>
</tr>
<tr>
<td>c13 Likert</td>
<td></td>
</tr>
<tr>
<td>c14 Likert</td>
<td></td>
</tr>
<tr>
<td>c15 Likert</td>
<td></td>
</tr>
<tr>
<td>c16 Likert</td>
<td></td>
</tr>
<tr>
<td><strong>Self-Discipline</strong></td>
<td></td>
</tr>
<tr>
<td>c17 TLTM</td>
<td></td>
</tr>
<tr>
<td>c18 TLTM</td>
<td></td>
</tr>
<tr>
<td>c19 TLTM</td>
<td></td>
</tr>
<tr>
<td>c20 TLTM</td>
<td></td>
</tr>
<tr>
<td><strong>Deliberation</strong></td>
<td></td>
</tr>
<tr>
<td>c21 TLTM</td>
<td></td>
</tr>
<tr>
<td>c22 TLTM</td>
<td></td>
</tr>
<tr>
<td>c23 TLTM</td>
<td></td>
</tr>
<tr>
<td>c24 TLTM</td>
<td></td>
</tr>
</tbody>
</table>

Note. $\chi^2 = 341.36$ (147), $p < .001$, CFI = .933, TLI = .875, RMSEA = .063. *p < .05, **p < .01, ***p < .001. Each item’s primary loading is highlighted in boldface.
Figure 1

Item-level associations between Likert and too little/too much (TLTM) ratings (Panel A) and their frequencies (Panel B) for Item 1 “I am neat and tidy” (C1: Order)