An Inventory of Problems–29 Study on Random Responding Using Experimental Feigners, Honest Controls, and Computer-Generated Data

(Article begins on next page)

14 December 2021

An Inventory of Problems – 29 (IOP-29) Study on Random Responding using Experimental Feigners, Honest Controls, and Computer-Generated Data

SCHOLARONE™
Manuscripts

ON RANDOM RESPONDING AND FEIGNING

Abstract

Self-reports may be affected by two primary sources of distortion, i.e., content-related (CRD) and content-unrelated (CUD) distortions. CRD and CUD, however, are often interconnected with each other, and similar detection strategies have been used to capture both. Thus, we hypothesized that a scale developed to detect random responding – arguably, one the most evident examples of CUD – would likely be sensitive to both CUD and, albeit to a lesser extent, CRD. Study 1 ($N = 1,901$) empirically tested this hypothesis by developing a random responding scale (RRS) for the recently introduced, Inventory of Problems – 29 (IOP-29; Viglione et al., 2017), and by testing it with both experimental feigners and honest controls. Study 2 ($N = 700$) then evaluated whether this newly developed RRS would perform similarly well with data from human participants instructed to respond at random versus computer-generated, random data. Interestingly, the sensitivity of our RRS dropped dramatically when considering the data from human participants. Together with the results of additional analyses inspecting the patterns of responses provided by our human random responders, these findings thus posed us a major question, i.e., is humans' random responding really random?

Keywords: IOP-29; Malingering; Random; Distortion; Response Style.

ON RANDOM RESPONDING AND FEIGNING

An Inventory of Problems – 29 (IOP-29) Study on Random Responding using

Experimental Feigners, Honest Controls, and Computer-Generated Data

A major problem of self-administered, psychological tests is that they typically rely on

the assumption that the test-taker is willing to and capable of responding to the test items in an

appropriate and truthful manner (Bornestein, 2015, 2017; Erard & Evans, 2017; Hopwood &

Bornstein, 2014; Meyer et al., 2001; Mihura, Meyer, Dumitrascu, & Bombel, 2013). If this

assumption is violated, for example when the examinee responds at random due to fatigue or

poor motivation, or when s/he makes an active effort to hide (e.g., in a child custody evaluation)

or to exaggerate (e.g., in a disability claim evaluation) his/her psychological weaknesses,

different types of response distortions occur. As an over-simplification of this multifaceted and

complex phenomenon, response distortions may be organized into two chief, broad categories:

content-related (CRD) and content-unrelated (CUD) distortions (Nichols, Greene, & Schmolck,

1989). The first (CRD) refers to situations in which the test-taker does not want to, or does not

know how to, respond appropriately or truthfully to a given item, because of its content. For

example, an individual prone to paranoia and persecutory ideas might intentionally choose to

answer "No" to an item inquiring whether s/he has ideas of persecution because s/he does not

want to reveal that s/he has that specific problem (e.g., in a screening for a high responsibility

job), or because s/he is genuinely unaware of it and truly thinks s/he has no persecutory ideas.

The second broad category of response distortions (CUD) is characterized by the fact that the

content of the item is not the reason why the examinee does not respond to it in an appropriate or

truthful way. For example, in long and complex personality inventories such as the Minnesota

Multiphasic Personality Inventory (MMPI-2; Green, 1991; MMPI-RF; Ben-Porath & Tellegen,

2

ON RANDOM RESPONDING AND FEIGNING

2008) or the Personality Assessment Inventory (PAI, Morey, 1996, 2007), the test-taker may at

some point get confused and place an answer in a wrong spot of the response sheet by distraction

error, or may get tired and respond to one or more items in a random, patterned (e.g., *T, F, T, F,*

...) or fixed (e.g., *T, T, T, T,* …) way, simply to finish faster.

The differentiation between CRD versus CUD, however, is more a theoretical rather than

a practical matter, as these two phenomena may co-occur and interact with each other, in real-life

assessment (Giromini et al., 2019). For instance, a poorly cooperative person inclined to

exaggerate or minimize a given psychopathological trait, when assessed with a long and complex

personality inventory such as the MMPI may at some point decide speed up the process by only

taking a quick glance at the items, rather than dedicating them the full attention they would

require. In this case, CUD would occur by the fact that part of test would be attended to with

poor attention, and CRD would co-occur by the fact that the content of the items attended to with

full attention would probably trigger some exaggeration or minimization. Additionally, CRD and

CUD might interact with each other, too. This would happen, for example, if while looking at the

items attended to with poor attention, the same person of the example above implicitly screened

for words relevant to the trait s/he is exaggerating or minimizing, and paid less attention to the

other words and items.

Not surprisingly, thus, the strategies adopted by researchers to detect CRD and CUD also

share a great deal of overlapping. In fact, for both types of distortion, two of the most widely

investigated approaches arguably involve searching for hints of *infrequent* and/or *inconsistent*

responding (Ben-Porath & Tellegen, 2008; Green, 1991; Fronczyk, 2014; Keeley et al., 2016;

Kelley et al., 2017; LePage, Mogge, & Garcia-Rea, 2009; Huang et al., 2015; Morey, 1996,

2007; Osborne & Blanchard, 2011; Pinsoneault, 2007; Rogers & Bender, 2018). In the case of

3

ON RANDOM RESPONDING AND FEIGNING

CRD (e.g., exaggeration, minimization, etc.), infrequent responses may reveal a tendency toward

endorsing uncommon weaknesses or virtues, and inconsistent responses may reveal an

unrealistic or untruthful symptom presentation characterized by contradictions. For example, a

person feigning depression might endorse dramatic complaints that are very infrequent among

true patients because s/he does not know what symptoms truly characterize that disorder (e.g., "I

never smiled in my life, not even once" – *True*), and s/he might respond inconsistently when

going from one item to another because of a proneness to respond in a more pathological

direction to all items (e.g., "When I go to bed, my problems get worse" – *True*; "When I wake

up, that is the worst part of my day" – *True*). In the case of CUD (e.g., random responding, fixed

response style, etc.), infrequent and inconsistent responding occur simply because the responses

given by the test-taker are independent from the content of the items, so that they do not follow

any specific logical patterns. As such, the likelihood of presenting contradictions and rare items

endorsement increases dramatically.

It should be noted, however, that the extent to which distortions conventionally assumed

to be content unrelated really are completely unaffected by the content of the items has not been

thoroughly investigated, so far. Indeed, Giromini et al. (2019) noted that most studies addressing

random responding – arguably one the most obvious examples of CUD – have been conducted

by comparing the responses given by human participants engaged in a testing situation, against

those produced by computer-based algorithms mimicking random responding (e.g., Archer &

Elkins, 1999; Fronczyk, 2014; Kelley et al., 2017; LePage et al., 2009). Occasionally, responses

deemed to be random responses have been obtained also by providing examinees with the test

response sheets only, but not with the booklets with the actual questions, so that they were blind

to the content of the items (e.g., Baer et al., 1999). To date, however, no studies have yet

4

ON RANDOM RESPONDING AND FEIGNING

researched whether the data produced by computer algorithms mimicking random responding truly resemble those produced by human participants instructed to respond at random. Along the same lines, the impact of having versus not having access to the test booklets, when examining data from human random responders, also is rather unknown. Because in real-world assessment, the test-taker is always a human being, who does have access to the text of the items s/he has to respond to, further research on this topic would certainly be beneficial. Indeed, if being exposed to the text of the test items implicitly influenced the patterns of endorsements provided by the so-called random responders, then the currently available, random responding scales (RRSs) might not be able to detect random responding as accurately as they have done in most of the currently published, scale development or cross-validation studies.

Overview of the Current Project

The original aim of the current project was to develop a scale to detect a major form of CUD, i.e., random responding, for use with a self-administered test that has been developed to evaluate a major source of CRD, i.e., negative response bias: The Inventory of Problems – 29 (IOP-29; Viglione, Giromini, & Landis, 2017). Briefly, the IOP-29 is a relatively new, 29-item test designed to assist practitioners evaluating the credibility of various psychiatric and cognitive symptom presentations. Twenty-seven of the 29 IOP-29 items offer three response options, i.e., *True*, *False*, and *Doesn't Make Sense*; the remaining two items are open-ended, cognitive items, calling for a numeric answer. Currently, the only scale embedded in the IOP-29 is the False Disorder Probability Score (FDS), a logistic regression-derived score reflecting the likelihood of obtaining a given IOP-29 from a sample of bona fide patients versus experimental feigners. While the FDS addresses CRD (more in detail, negative response bias), no IOP-29 scales have

ON RANDOM RESPONDING AND FEIGNING

yet been developed to evaluate CUD. This is the main reason why the current project was

initiated.

As noted above, wanting to develop a RRS, two major issues needed to be taken into

consideration: first, CUD and CRD may interact with each other, and similar strategies are often

adopted to detect both types of response distortion; second, it is currently unknown the extent to

which the patterns of responses provided by real-life random responders truly resemble those

coming from computer-generated, random data. The direct implication of the first of these

considerations is that a RRS based on patterns of endorsements that are infrequent and/or

inconsistent among well-motivated, honest responders (i.e., in the absence of CRD) would likely

end up being sensitive not only to random responding (i.e., CUD) but also to CRD (in particular,

in our case, negative response bias). Thus, to develop a "pure" RRS, i.e., a scale that detects

random responding only, rather than a mix of CUD and CRD, in addition to examining data from

honest responders, Study 1 also inspected data from experimental simulators, i.e., individuals

instructed to feign a cognitive or psychiatric disorder, via malingering experimental paradigm.

Our main hypothesis was that the patterns of endorsement that are idiosyncratic (infrequent)

and/or contradictory (inconsistent) for nonclinical or patient responders might characterize both

random responders and experimental simulators. Conversely, those that are idiosyncratic and/or

contradictory for experimental simulators, would be idiosyncratic and/or contradictory also for

honest responders, but not for random responders. Thus, we speculated that a pure RRS should

focus on hints of infrequency and inconsistency observed within experimental simulator samples,

rather than within honest or bona fide patients, so to pull apart from the scale any sources of

CRD.

6

ON RANDOM RESPONDING AND FEIGNING

Related to the fact that computer-generated, random data might be poorly representative

of IOP-29 responses obtained from actual human beings instructed to respond at random, we

decided that our scale needed to be tested with multiple data sets. As such, Study 2 inspected

four different sources of data: 1) a set of computer-generated, IOP-29s mimicking random

responding; 2) a set of IOP-29s obtained from human participants asked to take the IOP-29

without having access to the IOP-29 items; 3) a set of IOP-29s obtained from human participants

who did have access to the full IOP-29, but were instructed to respond at random; 4) a set of

IOP-29s from uncooperative respondents instructed to respond with a random-like approach,

mixing experimental feigning of psychopathology with random responding. This latter data set

was comprised of four subsamples instructed to feign four different mental health conditions, so

that Study 2 ultimately inspected 7 different subsets of data.

Study 1: Development and Initial Validation of a "Pure" Random Responding Scale

Study 1 aimed at developing a "pure" random responding scale (RRS), i.e., a scale that

would be independent from CRD. As noted above, because CRD and CUD are typically

identified by using the same detection strategies, we were concerned that a RRS developed with

data unaffected by CRD would likely end up measuring not only random responding, but also

CRD. Accordingly, Study 1 developed a first RRS based on the patterns of responses that were

infrequent and/or inconsistent among well-motivated, honest responders (Hon-RRS), and a

second RRS based on the patterns of responses that were infrequent and/or inconsistent among

experimental simulators (Sim-RRS). Our main hypothesis was that while both Hon-RRS and

Sim-RRS would be elevated by random responding, the former but not the latter would be

elevated also by experimental feigning. Said differently, we anticipated that Sim-RRS would be

7

ON RANDOM RESPONDING AND FEIGNING

the only one to measure pure random responding (i.e., CUD only), whereas Hon-RRS would

measure a mix of random responding and negative response bias (i.e., a mix of CUD and CRD).

Materials and Methods

Study 1 inspected archival IOP-29 data from a sample of honest respondents (bona fide

patients and nonclinical controls) and a sample of experimental simulators, and developed an

infrequency and an inconsistency indicators based on each of these samples. The infrequency

and inconsistency indicators derived from each sample were then combined with each other to

produce two RRSs, one derived from the honest responders sample (i.e., Hon-RRS) and one

derived from the experimental simulators sample (i.e., Sim-RRS). Lastly, the Hon-RRS and Sim-

RRS were tested on three independent data sets including honest responders, experimental

simulators, and computer-generated IOP-29 data mimicking random responding.

Participants. Three sets of IOP-29 data were used for this study. A first data set included

891 archival data from clinical and nonclinical honest responders; a second data set included 910

archival data from experimental simulators; a third data set included 100 computer data

generated ad hoc for this study to mimic random responding.

Honest Responders. An archival data set of 891 IOP-29s from honest responders,

including 491 bona fide patients and 400 nonclinical controls was used to develop and validate

an indicator of infrequency and an indicator of inconsistency based on responding patterns of

honest test-takers. Of these 891 IOP-29s, 275 came from the US (Viglione et al., 2017) and 616

from Italy (216 from Giromini et al., 2018; 400 from Giromini et al., 2019). The patient

participants from Viglione et al. (2017) and Giromini et al. (2018) had diagnoses of mild

traumatic brain injury-, depression-, anxiety-, psychosis-, and trauma-related disorders. As for

the 400 healthy respondents from Giromini et al. (2019), they were Italian volunteers who did

8

ON RANDOM RESPONDING AND FEIGNING

not suffer from any specific mental health problems. In the original study, these 400 research

participants were instructed to take the IOP-29 three times, in three different conditions. In one

condition, they were asked to respond honestly; in one condition, they were asked to feign a

mental health disorder via malingering experimental paradigm; in one condition, they were asked

to respond as if they were mentally ill and wanted to respond with an uncooperative or random-

like approach. The honest responders subsample of the current study only includes the 400 IOP-

29s coming from the first of these three conditions.

In terms of demographics, the combined sample encompassing 891 adult volunteers was

highly heterogeneous regarding gender, age, education, racial characterization, and marital status

(for details, see Giromini et al., 2018; Giromini et al., 2019; Viglione et al., 2017). The great

majority of this sample, i.e., $n = 791$, was used to develop the Hon-RRS (Honest Responders –

Developmental Sample); 100 were randomly extracted to be used for cross-validation purposes

(Honest Responders – Validation Sample).

Experimental Simulators. As for the sample of experimental simulators, an archival data

set of 910 IOP-29s, 274 from the US (Viglione et al., 2017) and 636 from Italy (236 from

Giromini et al., 2018; 400 from Giromini et al., 2019), contributed to this research. As for the

400 IOP-29s from Giromini et al. (2019), they were taken from the second condition described

above, in which participants were asked to take the test as if they wanted to convince the

examiner that they were mentally ill. In all cases, simulators were provided with a brief scenario

or vignette from real-life situations, and were warned not to be too dramatic in their symptom

presentations, so to improve their ability to feign in a more credible way (Rogers & Bender,

2013). In terms of conditions to be feigned, each participant was instructed to pretend to suffer

9

ON RANDOM RESPONDING AND FEIGNING

from one or more of the following: cognitive/neuropsychological impairment, psychosis, PTSD, and depression.

Also the combined, experimental simulators sample was highly heterogeneous regarding gender, age, education, racial characterization, and marital status (for details, see Giromini et al., 2018; Giromini et al., 2019; Viglione et al., 2017). These data served to develop and validate an indicator of infrequency and an indicator of inconsistency based on responding patterns of experimental simulators. Consistent with the procedures followed for the honest responders sample, the majority of these data, i.e., $n = 810$, were used to develop the Sim-RRS (Experimental Simulators – Developmental Sample), whereas 100 IOP-29s were used for cross-validation purposes (Experimental Simulators – Validation Sample).

Computer Generated Data. To test the sensitivity of the RRSs developed based on the honest responders' and experimental simulators' response patterns, a set of 100 computer generated, IOP-29 data was used. These IOP-29s were created with the goal of mimicking pure random responding. Thus, for the 27 IOP-29 items that offer three response options (i.e., *True*, *False*, and *Doesn't Make Sense*), responses were generated with a computerized function to create random data and by setting, per each item, a probability of 33.33% per each response option. For the two cognitive, open-ended items, as they call for a numeric answer and the size of the space dedicated to them in the response sheets implicitly suggests a maximum of three digits, we assigned to each an integer number from 0 to 999 with equal probability for each number in that range.

Data Analysis. Data analysis initially focused on the developmental samples comprised of 791 honest responders and 810 experimental simulators, aiming at identifying infrequent and/or inconsistent patterns of endorsement within each of these two groups. As for the

10

ON RANDOM RESPONDING AND FEIGNING

infrequency indicators, an item-response combination (e.g., *True* to Item #17) was selected to be part of the infrequency indicator if its frequency was lower than 15% in the sample under consideration. As for the inconsistency indicators, a pair of item-response combinations (e.g., *False* to Item #2 & *True* to Item #6) was selected to be part of the inconsistency indicator if the two item-response combinations of that pair correlated with each other at least *r* = .30 in absolute value. For example, if the responses *False* to Item #9 and *True* to Item #18 correlated with each other with *r* ≥ .30, then responding *False* to Item #9 and *False* to Item #18, or *True* to Item #9 and *True* to Item #18 would be considered to be an index of inconsistent responding.

The item-response combinations selected for the infrequency scales and the pairs of item-response combinations selected for the inconsistency scales were then assigned one point each, and the values obtained by summing up these points represented the scores of the final, infrequency and inconsistency indicators under investigation. Next, the Hon-RRS was calculated as the z sum of the infrequency and inconsistency indicators developed using the sample of honest responders, and the Sim-RRS was calculated as the z sum of the infrequency and inconsistency indicators developed using the sample of experimental simulators. Both scales were then converted into T scores, to facilitate their readability and interpretation (Giromini et al., 2017; Meyer et al., 2007).

Lastly, the validity of the Hon-RRS and Sim-RRS was tested with the independent validation samples including 100 IOP-29s from honest responders, 100 IOP-29s from experimental simulators, and 100 IOP-29s generated via computer to mimic random responding. More specifically, two one-way ANOVAs were performed, with data source as the independent variable, and the Hon-RRS and Sim-RRS scores as the dependent variables. Furthermore, the

ON RANDOM RESPONDING AND FEIGNING

sensitivity and specificity of the Hon-RRS and Sim-RRS were examined too, by considering two frequently inspected cut-off scores, i.e., $T \geq 65$ and $T \geq 70$ (Giromini et al., 2017).

Results

Development of the RRSs. Within the honest responders developmental sample, 28 item-response combinations had a frequency of endorsement lower than 15%, and 18 pairs of item-response combinations yielded correlation values equal to or greater than $r = .30$ in absolute value. Accordingly, the infrequency and inconsistency scales developed based on the honest responders sample included 28 item-response combinations and 18 pairs of item-response combinations, respectively. Within the experimental simulators developmental sample, the number of item-response combinations selected for the infrequency indicator was 25, and the number of pairs of item-response combinations selected for the inconsistency scale was 33.

By combining these newly developed, infrequency and inconsistency indicators within each developmental sample, two composite, RRSs were created. As noted above, Hon-RRS was obtained by combining via z sum the infrequency and inconsistency indicators developed using the honest responders developmental sample; Sim-RRS was obtained by combining via z sum the infrequency and inconsistency indicators developed using the experimental simulators developmental sample. To facilitate readability and interpretation, both Hon-RRS and Sim-RRS were then converted into T scores based on the mean and standard deviation values observed in their respective, developmental samples.

Our main hypothesis was that the RRS based on the patterns of infrequent and/or inconsistent responding found within a sample of honest responders would measure both CUD and CRD, whereas the RRS developed with the experimental simulators sample would only measure CUD. In line this hypothesis, Hon-RRS was significantly higher within the

ON RANDOM RESPONDING AND FEIGNING

experimental simulators sample (M = 71.9; SD = 22.5) than within the honest responders sample

(M = 50.0; SD = 14.4), $t(1378.9)$ = 23.3, p < .01, d = 1.16,[1] whereas Sim-RRS did not

statistically differ from one sample to another, $t(1535.8)$ = 1.5, p = .14, d = .07[1] (experimental

simulators had a mean of 50.0, SD = 15.7; honest responders had a mean of 51.0; SD = 12.5).

Also consistent with our expectations, Hon-RRS positively correlated with the FDS of the IOP-

29 (i.e., the chief, negative response bias score of the IOP-29; Viglione et al., 2017), with large

effect sizes (Cohen, 1988), whereas Sim-RRS produced only small and/or nonsignificant

correlations (Table 1). Also noteworthy, Hon-RRS and Sim-RRS positively correlated with each

other, which suggests that both scores probably share a common variance, presumably associated

with CUD.

      Initial Validation of the RRSs. To provide a first, independent, cross-validation for

these newly developed RRSs, three independent, validation samples encompassing 100 honest

responders, 100 experimental simulators and 100 computer-generated data mimicking random

responding were inspected. As shown in Table 2, the scores of Hon-RRS were notably different

across the three samples, $F(2, 297)$ = 203.4, p < .001, with all Bonferroni corrected post hoc tests

being statistically significant at p < .001. The highest Hon-RRS scores were observed for the

computer-generated data, followed by the experimental simulators data, and lastly by the honest

responders data. For Sim-RRS, the main effect was statistically significant too, $F(2, 297)$ =

187.0, p < .001. In this case, however, Bonferroni corrected post hoc tests revealed that the

scores of honest responders did not significantly differ from those of experimental simulators, p

≈ 1.00, and that both these groups scored significantly lower compared to the computer-

---

[1] Because homoscedasticity could not be assumed, Welch-Satterthwaite method was used to adjust degrees of freedom.

13

ON RANDOM RESPONDING AND FEIGNING

generated data set, *p* < .001. These data provide additional, and perhaps more stringent evidence

that Hon-RRS measures both CUD and CRD, whereas Sim-RRS measures CUD only.

Lastly, Table 3 presents information concerning the sensitivity and specificity of Hon-

RRS and Sim-RRS, based on two commonly investigated cut-off points for T scores, i.e., T ≥ 65

and T ≥ 70 (Giromini et al., 2017). In line with the hypothesis that both scales are sensitive to

CUD, when considering the computer-generated data, both yielded very high sensitivity values,

ranging from .89 to .95. Nonetheless, for Hon-RRS, specificity ranged from .85 to .90 when

considering the honest responders sample, but dropped to .44 to .56 when considering the

experimental simulators sample. Conversely, Sim-RRS was similarly specific in both the

samples, with specificity values ranging from.89 to .92 in the honest responders sample, and

from .86 to .90 in the experimental simulators sample.

Discussion

Because CRD and CUD often co-occur and interact with each other, and are typically

identified by using the same detection strategies, Study 1 was designed to test whether

developing a RRS based on the patterns of infrequent and/or inconsistent responding observed in

a sample of experimental simulators would allow us to pull apart CRD-related variance from

CUD-related variance. That is, we speculated that a RRS developed using a sample of honest

responders would end up measuring both CUD and CRD, whereas a RRS developed using a

sample of experimental simulators would measure CUD only. Our results fully confirmed this

hypothesis: While random responding elevated both Hon-RRS and Sim-RRS, negative response

bias only elevated Hon-RRS, without affecting Sim-RRS. Sim-RRS may thus be preferable over

Hon-RRS, as it is similarly sensitive to pure random responding, but less affected by other

sources of distortions such as negative response bias.

14

ON RANDOM RESPONDING AND FEIGNING

Study 2: Cross-Validation with PC versus Human Participants

Study 2 aimed at testing the extent to which the findings observed with computer-generated IOP-29 data mimicking random responding would generalize to those obtained with human participants instructed to respond at random or in an uncooperative way. Additionally, it also sought to investigate whether having versus not having access to the text of the test items would influence the patterns of endorsement of human participants instructed to respond at random. Sim-RRS obtained from Study 1 was thus tested in Study 2 with three additional data sets: 1) a set of newly collected IOP-29s obtained from human participants asked to take the IOP-29 without having access to the IOP-29 items; 2) an archival set of IOP-29s obtained from human participants who did have access to the full IOP-29, but were instructed to respond at random; 3) a set of archival, IOP-29s from uncooperative respondents instructed to respond with a random-like approach, mixing experimental feigning of psychopathology with random responding. As anticipated above in the Introduction, this latter data set was comprised of four subsamples instructed to respond with a random-like approach to pretend they suffered from four different kinds of mental health problems.

Materials and Methods

All human participants included in Study 2 were nonclinical volunteers from Italy, who had signed an informed consent form prior to being enrolled in the study. Five-hundred came from archival research (Giromini et al., 2019), whereas 100 were newly collected for this study. Additionally, the computer-generated data set from Study 1 including 100 IOP-29s mimicking random responding was re-utilized too. Accordingly, the total sample size for Study 2 is N = 700.

ON RANDOM RESPONDING AND FEIGNING

Participants. Seven different data sets, or subsamples, were included in the analyses of

Study 2, as detailed below. Each subsample included 100 IOP-29s. The last five subsamples, i.e.,

subsamples 3, 4, 5, 6, and 7, came from the same archival data set, by Giromini et al. (2019).

Subsample 1 (S1), Computer-Generated Data (from Study 1). S1 refers to the 100

computer-generated IOP-29s from Study 1. This sample was used to evaluate the extent to which

the validity of Sim-RRS would generalize from computer-generated data to data from human

participants instructed to respond at random or in an uncooperative way. Details about the

algorithm to generate random IOP-29 responses are reported above (see Study 1).

Subsample 2 (S2), Blind Random Responders (newly collected data). S2 is comprised of

100 Italian adult volunteers, instructed to respond at random to the IOP-29 without having access

to the IOP-29 items, i.e., blind to the actual content of the test items. More in detail, participants

were instructed that their task would be to fill out the IOP-29 by responding at random; then they

were given the IOP-29, with the text of the items covered by a piece of paper, so that they could

only see the IOP-29 instructions and the response options. In terms of demographic information,

58% were men, ages ranged from 18 to 68 ($M = 37.1$, $SD = 13.2$), about half (i.e., 56%) were not

in a relationship, and a little less than half had a college degree or more (i.e., 40%).

Subsample 3 (S3), Real-Life Random Responders (from Giromini et al., 2019). S3

includes a set of 100 IOP-29s, which has been previously used by Giromini et al. (2019) to test

the impact of random responding on IOP-29 FDS scores (section entitled "Additional

Analyses"). These 100 IOP-29s were obtained from a sample of 100 Italian volunteers who had

been handed the IOP-29 in its standard form, i.e., without covering the text of the items of the

test, but with the instruction to respond at random. The procedures used to collect these data are

fairly similar to those used in real-life assessment, except for the request to respond at random.

16

ON RANDOM RESPONDING AND FEIGNING

As such, S3 is highly suitable to determine the extent to which computer-generated, random data sets may resemble random responses obtained in real-life assessment. In terms of demographic information, 54% were men, ages ranged from 18 to 46 ($M = 24.4$, $SD = 4.5$), the great majority (i.e., 90%) were not in a relationship, and a little more than half had a college degree or more (i.e., 64%).

Subsample 4 (S4), Uncooperative NP (from Giromini et al., 2019). S4 is a subsample extracted from another, relatively large, archival data set used by Giromini et al. (2019). As noted above, these research participants were instructed to take the IOP-29 three times, in three different conditions. Subsamples 4, 5, 6, and 7 of the current study include the IOP-29s taken from the condition in which participants were instructed to respond in an uncooperative or resistant way, without paying attention to the content of the items, and by selecting their responses at random, so to not truly reveal themselves to the examiner. More specifically, they were asked to respond as if they were mentally ill and did not want to open up to the examiner as they had applied to receive a monetary compensation, and thought that undergoing psychological testing could undermine their likelihood to obtain such compensation, so that they eventually decided to respond with this random-like approach. Four subsamples were included in Giromini et al. (2019), based on what condition they were asked to pretend they suffered from: 1) neuropsychological problems following traumatic brain injury (NP); 2) schizophrenia-related problems (SCZ); 3) PTSD-related problems (PTSD); 4) depression-related problems (DEP). S4 includes the 100 Italian participants who had been instructed to take the IOP-29 as if they were suffering from neuropsychological impairment following mild traumatic brain injury (NP) and did not want to reveal themselves to the examiner, so that they decided to respond in an uncooperative way, using a random or random-like approach. In terms demographic information,

17

ON RANDOM RESPONDING AND FEIGNING

36% were men, ages ranged from 19 to 70 (M = 40.5, SD = 14.7), a little more than half (i.e.,

55%) were not in a relationship, and a little more than half had a college degree or more (i.e.,

65%).

   Subsample 5 (S5), Uncooperative SCZ (from Giromini et al., 2019). S5 also comes from

Giromini et al. (2019). In this case, participants (100 Italian adults) were given the description of

a person affected by schizophrenia (SCZ), who decided to respond at random, or in an

uncooperative way. In this subsample, 50% were men, ages ranged from 18 to 68 (M = 39.4, SD

= 15.8), a little more than about half (i.e., 57%) were not in a relationship, and about a third had a

college degree or more (i.e., 29%).

   Subsample 6 (S6), Uncooperative PTSD (from Giromini et al., 2019). Like S4 and S5,

this subsample also comes from Giromini et al. (2019). In this case, participants (100 Italian

adults) were asked to respond at random or in an uncooperative way because they suffered from

PTSD symptoms and did not want to risk their right to receive a monetary compensation that

they had applied for. In terms demographic information, 44% were men, ages ranged from 18 to

63 (M = 43.6, SD = 10.9), about a third (i.e., 32%) were not in a relationship, and about a quarter

had a college degree or more (i.e., 24%).

   Subsample 7 (S7), Uncooperative DEP (from Giromini et al., 2019). Like S4, S5, and

S6, also S7 was taken from Giromini et al. (2019). Differently from the other subsamples, the

reason why participants in this subsample (100 Italian adults) decided to respond with a random-

like or uncooperative approach is that they had applied for a monetary compensation related to

their being currently affected by major depression. In this subsample, 36% were men, ages

ranged from 18 to 70 (M = 41.6, SD = 15.2), about half (i.e., 49%) were not in a relationship, and

a little less than half had a college degree or more (i.e., 41%).

ON RANDOM RESPONDING AND FEIGNING

Data Analysis. The major goal of Study 2 was to evaluate the extent to which Sim-RRS scores would vary from one subsample to another. Accordingly, a one-way ANOVA was performed, with subsample as the independent variable, and Sim-RRS as the dependent variable. Next, similarly to Study 1, the sensitivity of Sim-RRS to detect random responding was examined by using $T \geq 65$ and $T \geq 70$ as cut-off values.

Results

The average Sim-RRS scores for each of the seven subsamples of Study 2 are presented in Table 4. A graphical representation of these same data is also available, in Figure 1. The main effect of data source was statistically significant, $F(6, 693) = 17.5$, $p < .001$. Interestingly, Bonferroni corrected, post hoc analyses revealed that S1 yielded significantly higher scores compared to all other six subsamples, $p < .001$; S2 yielded significantly lower scores than S1, but significantly higher scores compared to all other five subsamples, $p < .001$; the scores from S3, S4, S5, S6, and S7 did not significantly differ from each other, $p \approx 1.00$.

Examination of sensitivity results, reported in Table 5, also leads to similar considerations. Regardless of what cut-off score is taken under investigation, the highest sensitivity values were found for S1 (.93 for $T \geq 65$ and .89 for $T \geq 70$), followed by S2 (.83 for $T \geq 65$ and .68 for $T \geq 70$), and then by all other subsamples, with small variations from one to another ($\leq .56$ for $T \geq 65$ and $\leq .49$ for $T \geq 70$). Thus, Sim-RRS achieved excellent sensitivity with S1, good sensitivity with S2, and much less satisfactory sensitivity with all other subsamples.

Additional Analyses. Because the results of Study 2 revealed that the computer-generated data were significantly different from all other sources of data from human participants instructed to respond at random or with a random-like approach, we decided to

19

ON RANDOM RESPONDING AND FEIGNING

perform additional analyses aimed at better understanding this intriguing pattern of findings.

More specifically, we performed a series of Chi^2 statistics, item by item, to evaluate the extent to

which the responses provided by our human participants did versus did not meet a-priori

expectations for pure random responding. These analyses excluded the IOP-29 items that offer

open-ended response options, as the expected response frequency for each of their theoretically

infinite response options would approximate zero. For all remaining items, given that they offer

three response options (i.e., *True*, *False*, and *Doesn't Make Sense*), a probability of 33.33% was

assigned to each response option. Indeed, if the participants responded completely at random, a

frequency of 33.33% per each response option should be expected.

The results of these additional analyses, summarized in Table 6, are quite surprising. In

most cases, despite the instruction to respond at random or with a random-like approach, our

participants did not really respond completely at random. For example, both the blind random

responders of S2 and the real-life random responders of S3 responded *True* to Item 1 with an

extremely high frequency, approximating 70%. Also noteworthy, the response option *Doesn't

Make Sense* was endorsed relatively rarely by all six subsamples. Indeed, when considering the

combined sample comprised of 600 human participants instructed to respond at random or with a

random like approach, the frequency of endorsement of *Doesn't Make Sense* ranged from 8% to

29%, with an average frequency of 20% across all 27 items under consideration.

Finally, we also examined possible patterns of fixed responding. More specifically, we

examined the total number of responses in which a given item presented the same response

option endorsed also for the previous item (e.g., *True* to Item #24 after endorsing *True* to Item

#23), and compared this number across all seven Study 2 subsamples. The main effect of the

one-way ANOVA testing possible differences from one subsample to another, however, was not

20

ON RANDOM RESPONDING AND FEIGNING

statistically significant, $F(6, 639) = 1.99$, $p = .065$. More importantly, Bonferroni corrected post

hoc comparisons revealed no significant differences between the computer-generated data set

and any of the six subsamples encompassing human participants, $p \geq .692$. These findings thus

suggest that fixed responding is not the reason why our human participants instructed to respond

at random or with a random-like approach did not really respond following pure random

responding expectations.

Discussion

The primary purpose of Study 2 was to test the extent to which the findings concerning

the sensitivity of a RRS obtained with computer-generated data would replicate to data obtained

with human participants instructed to respond at random or with a random-like approach.

Additionally, Study 2 also aimed at comparing the findings obtained with human random

responders who did versus did not have access to the text of the test items. Taken together, the

results of Study 2 indicate that using computer-generated data or – albeit to a lesser extent – data

coming from human participants blind to the text of the test items may artificially boost the

sensitivity estimates of a RRS, compared to using data collected by administering the test in its

standard format and instructing participants to respond at random.

General Discussion and Final Remarks

The two studies described in this paper contribute to better understanding the

phenomenon of random responding in two primary ways. First, Study 1 demonstrates that the

manifestations of CUD and CRD are so closely interconnected with each other, that if a RRS is

developed based on response patterns unaffected by CRD, then that RRS will likely end up

measuring both CRD and CUD. To pull apart content related from content unrelated variance, a

suitable approach might be to derive that RRS based on data known to be characterized by CRD.

21

ON RANDOM RESPONDING AND FEIGNING

Indeed, in our Study 1, the RRS developed based on the patterns of responses provided by honest responders (i.e., Hon-RRS) was sensitive to random responding but not specific to it, whereas the RRS developed based on the patterns of responses provided by experimental simulators (i.e., Sim-RRS) achieved similar sensitivity but greater specificity.

The second – and in our opinion most important – contribution of this article concerns the major warning emerging from Study 2: Researchers willing to develop and/or validate a RRS for any given, psychological assessment tools should avoid testing their scale(s) solely with computer-generated data mimicking random responding or with human random responders who do not have access to the text of the test items. Indeed, our results from Study 2 show that the sensitivity of our newly developed, Sim-RRS dropped dramatically when moving from a computer-generated data set mimicking random responding to a sample of human participants instructed to respond at random without having access to the text of the test items, and that it further decreased when moving from these two samples to a more ecologically valid sample of human participants administered the IOP-29 in its standard format, with the instruction to respond at random or with a random-like approach. Given that using computer-generated data (e.g., Archer & Elkins, 1999; Fronczyk, 2014; Kelley et al., 2017; LePage et al., 2009) or participants not having access to the items of the test (e.g., Baer et al., 1999) is a widely adopted practice in random responding-related investigations, our findings raise important questions concerning the actual sensitivity of many of the RRSs currently in use.

In terms of implications for IOP-29 research, Sim-RRS demonstrated excellent specificity but unsatisfactory sensitivity, when data other than that generated via computer was taken into consideration. As such, additional refinements to this scale are recommended. Future research, in particular, might attempt to investigate reaction times, given that random responding

ON RANDOM RESPONDING AND FEIGNING

might likely associate with infrequent and/or inconsistent reaction time patterns (Viglione et al., 2018). Additionally, future research could also focus on patterns of endorsement typical of random responders. Indeed, Study 2 showed that despite the instruction to respond at random, the great majority of test-takers did not truly respond completely at random. For example, almost 70% of the subsamples S2 and S3 endorsed *True* to Item #1, and 60% of S6 endorsed *False* to Item#16. Thus, understanding what patterns of endorsement are typical of human participants instructed to respond at random might help shedding some light on how to further refine and possibly improve the utility of Sim-RRS.

In particular, future research might attempt to replicate our Study 2 findings by changing the format with which the IOP-29 is presented. For example, it would be interesting to know whether Item #1 would be answered *True* by the 60% to 70% of the random responders also if it was presented at a later position, rather than as the first item. Along similar lines, it would be useful to evaluate whether the order with which the response options are presented also could affect the response patterns endorsed by the so-called random responders. For example, should the response option *Doesn't Make Sense* be the first response option rather than the third one, would it be chosen more frequently than what observed in our Study 2? Also, would the majority of the sample continue to endorse *True* to Item #1 should the order of the response options be reversed? Finding the answers to these questions would probably notably broaden our knowledge and understanding of the real mechanisms beyond the patterns of endorsement shown by human participants instructed to respond at random.

A few limitations should be kept in mind, however, when considering the implications of our findings. First, we only administered the IOP-29, which poses questions concerning the generalizability of our findings to other and different types of tests. For example, the big drop in

ON RANDOM RESPONDING AND FEIGNING

the sensitivity estimates observed in our Study 2 when moving from the computer-generated

subsample to the subsamples comprised of human participants instructed to respond at random

may not be so evident when inspecting longer and more complex personality assessment

inventories such as the MMPI or PAI. Indeed, it is possible that our participants did not respond

completely at random simply because the IOP-29 is a brief test, so that they were somehow

curious to take a look at the items and did not completely ignore their content, like we requested

them to do. This may not happen with a test like the MMPI or PAI. Besides, as we focused on 29

items only, we did not have as many data point to use to detect random responding as it occurs

with a longer and more complex test, meaning that our RRS may not be as precise and accurate

as the RRSs developed for other, longer and more complex tests. Furthermore, given that the

IOP-29 has been designed to address negative response bias, one might question whether the

findings of Study 1 would replicate also when using other tests, unrelated to CRD. For instance,

if rather than using the IOP-29 one used another test, unrelated to negative response bias (e.g.,

the Difficulties in Emotion Regulation Scale; Gratz & Roemer, 2004), would a RRS developed

based on infrequent and/or inconsistent responding patterns observed among honest responders

end up measuring both CUD and CRD like our Hon-RRS did? Lastly, another potential

limitation of our investigation is that our human random responders were all nonclinical adult

volunteers from Italy. Future studies might thus try to replicate our studies by involving

individuals from different countries and perhaps also individuals affected by psychopathology.

Despite these and potentially other limitations, the current study still has the merit to

provide a tentative approach to pull apart CRD from CUD-related indicators, and to pose a major

question to all researchers interested in random responding, i.e., is random responding really

random?

24

ON RANDOM RESPONDING AND FEIGNING

References

Archer, R. P., & Elkins, D. E. (1999). Identification of random responding on the MMPI-A.

Journal of Personality Assessment, 73, 407 –421

Baer, R. A., Kroll, L. S., Rinaldo, J., & Ballenger, J. (1999). Detecting and discriminating

between random responding and overreporting on the MMPI-A. *Journal of Personality*

Assessment, 72(2), 308-320.

Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF (Minnesota Multiphasic Personality*

Inventory-2 Restructured Form): Manual for administration, scoring, and interpretation.

Minneapolis. University of Minnesota Press.

Bornstein, R. F. (2015). Personality assessment in the diagnostic manuals: On mindfulness,

multiple methods, and test score discontinuities. *Journal of Personality Assessment, 97*,

446-455.

Bornstein, R. F. (2017). Evidence-based assessment of interpersonal dependency. *Journal of*

Personality Assessment, 91, 1-8.

Erard, R. E., & Evans, F. B. (2017). *The Rorschach in multimethod forensic assessment*

conceptual foundations and practical applications. New York, NY: US:

Routledge/Taylor & Francis Group.

Fronczyk, K. (2014). The identification of random or careless responding in questionnaires: The

example of the NEO-FFI. *Roczniki Psychologiczne, 17(2)*, 459-473.

Giromini, L., Ales, F., de Campora, G., Zennaro, A., & Pignolo, C. (2017). Developing age and

gender adjusted normative reference values for the Difficulties in Emotion Regulation

Scale (DERS). *Journal of Psychopathology and Behavioral Assessment, 39(4)*, 705-714.

Doi:10.1007/s10862-017-9611-0

ON RANDOM RESPONDING AND FEIGNING

Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2019). An Inventory of Problems – 29 (IOP-29) sensitivity study investigating feigning of four different symptom presentations via malingering experimental paradigm. *Manuscript accepted for publication in Journal of Personality Assessment.*

Giromini, L., Viglione, D.J., Pignolo, C. & Zennaro, A. (2018). A clinical comparison, simulation study testing the validity of SIMS and IOP-29 with an Italian sample. *Psychological Injury and Law, advance online publication*, doi:10.1007/s12207-018-9314-1

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment, 26(1)*, 41– 54.

Green, R.L. (1991). *The MMPI-2/MMPI: An interpretative manual*. Boston, MA: Allyn & Bacon.

Hopwood, C. J., & Bornstein, R. F. (2014). *Multimethod clinical assessment*. New York: Guildford.

Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2015). Detecting insufficient effort responding with an infrequency scale: evaluating validity and participant reactions. *Journal of Business and Psychology, 30*, 299-311.

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

ON RANDOM RESPONDING AND FEIGNING

Kelley, S. E., Edens, J. F., Donnellan, M. B., Somma, A., Fossati, A. , de Ruiter, C., Eisenbarth,

　　　H., & Vaughn, M. G. (2017). Detecting inconsistent responding on the Youth

　　　Psychopathic Traits Inventory–Short Form. *Assessment, Advance Online Publication*.

LePage, J., Mogge, N. L., Garcia-Rea, E. A. (2009). Detecting random responding using the

　　　Assessment of Depression Inventory: a brief screening measure of depression.

　　　Depression and Anxiety, 26, 592-595.

Meyer, G. J., Erdberg, P., Shaffer, T. W. (2007). Toward international normative reference data

　　　for the Comprehensive System. *Journal of Personality Assessment, 89(S1),* S201-S216.

Meyer, G. J., Finn, S. E., Eyde, L., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J.,

　　　Kubiszyn, T. W., & Reed, G. M. (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

Mihura, J. L., Meyer, G. J., Dumitrascu, N., & Bombel, G. (2013). The validity of individual

　　　Rorschach variables: Systematic reviews and meta-analyses of the Comprehensive

　　　System. *Psychological Bulletin, 139*, 548–605.

Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory (PAI)*.

　　　Odessa, FL: Psychological Assessment Resources.

Morey, L. C. (2007). *Personality Assessment Inventory (PAI). Professional Manual (2nd ed.)*.

　　　Odessa, FL: Psychological Assessment Resources.

Nichols, D. S., Greene, R. L., Schmolck, P. (1989). Criteria for assessing inconsistent patterns of

　　　item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of

　　　Clinical Psychology, 45(2)*, 239-250.

ON RANDOM RESPONDING AND FEIGNING

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*, 220.

Pinsoneault, T. B. (2007). Detecting random, partially random, and nonrandom Minnesota

Multiphasic Personality Inventory-2 protocols. *Psychological Assessment, 19(1)*, 159-164

Rogers, R. & Bender, D. (2018). *Clinical assessment of malingering and deception*. New York,

NY: Guilford Press.

Viglione, D. J., Giromini, L. & Landis, P. (2017). The development of the Inventory of

Problems–29: A brief self-administered measure for discriminating bona fide from

feigned psychiatric and cognitive complaints. *Journal of Personality Assessment, 99*,

534-544.

Viglione, D. J., Giromini, L., Landis, P., McCullaugh, J. M., Pizitz, T. D., O'Brien, S., Wood, S.,

Connell, K., & Abramsky, A. (2018). Development and Validation of the False Disorder

Score: The Focal Scale of the Inventory of Problems. *Journal of Personality Assessment,*

in press.

ON RANDOM RESPONDING AND FEIGNING

Table 1. Correlations between Hon-RRS, Sim-RRS, and IOP-29 FDS.

	IOP-29 FDS	Hon-RRS
Honest Responders – Developmental Sample (*n* = 791)		
Hon-RRS	.39**	
Sim-RRS	.14**	.65**
Experimental Simulators – Developmental Sample (*n* = 810)		
Hon-RRS	.45**	
Sim-RRS	-.11**	.47**
Combined, Developmental Sample (*n* = 1601)		
Hon-RRS	.62**	
Sim-RRS	-.04	.44**

* *p* < .05; ** *p* < .01.

ON RANDOM RESPONDING AND FEIGNING

Table 2. Comparison of Hon-RRS and Sim-RRS Scores (Study 1).

	Min	Max	M	SD
Honest Responders – Validation Sample (*n* = 100)				
Hon-RRS	25.2	105.3	50.5	14.7
Sim-RRS	18.3	90.5	50.1	12.8
Experimental Simulators – Validation Sample (*n* = 100)				
Hon-RRS	27.5	122.9	68.1	21.3
Sim-RRS	18.3	89.8	49.1	15.1
Computer-generated, Random Data (*n* = 100)				
Hon-RRS	40.5	138.4	100.8	17.2
Sim-RRS	36.1	118.3	83.3	14.6

ON RANDOM RESPONDING AND FEIGNING

Table 3. Sensitivity and Specificity of Hon-RRS and Sim-RRS (Study 1).

	Hon-RRS	Sim-RRS
Honest Responders – Validation Sample ($n = 100$)		
T ≥ 65 / T < 65	15 / **85**	11 / **89**
T ≥ 70 / T < 70	10 / **90**	8 / **92**
Experimental Simulators – Validation Sample ($n = 100$)		
T ≥ 65 / T < 65	56 / **44**	14 / **86**
T ≥ 70 / T < 70	44 / **56**	10 / **90**
Computer-generated, Random Data ($n = 100$)		
T ≥ 65 / T < 65	**95** / 5	**93** / 7
T ≥ 70 / T < 70	**94** / 6	**89** / 11

Note. Correct classifications are highlighted by bold format.

ON RANDOM RESPONDING AND FEIGNING

Table 4. Descriptive Statistics for Sim-RRS Scores across Study 2 Subsamples (N = 700).

	n	*Min*	*Max*	*M*	*SD*
Subsample 1 (S1), Computer-Generated Data	100	36.1	118.3	83.3	14.6
Subsample 2 (S2), Blind Random Responders	100	23.2	115.5	75.9	13.7
Subsample 3 (S3), Real-Life Random Responders	100	31.0	98.9	68.2	14.0
Subsample 4 (S4), Uncooperative NP	100	25.2	106.5	63.5	18.4
Subsample 5 (S5), Uncooperative SCZ	100	29.2	117.1	67.0	18.1
Subsample 6 (S6), Uncooperative PTSD	100	18.3	123.4	64.6	18.6
Subsample 7 (S7), Uncooperative DEP	100	18.3	118.3	67.0	20.3

ON RANDOM RESPONDING AND FEIGNING

Table 5. Sensitivity of Sim-RRS across Study 2 Subsamples (N = 700).

	Sim-RRS	
	$T \geq 65$ / $T < 65$	$T \geq 70$ / $T < 70$
Subsample 1 (S1), Computer-Generated Data	**93** / 7	**89** / 11
Subsample 2 (S2), Blind Random Responders	**83** / 17	**68** / 32
Subsample 3 (S3), Real-Life Random Responders	**51** / 49	**45** / 55
Subsample 4 (S4), Uncooperative NP	**47** / 53	**37** / 63
Subsample 5 (S5), Uncooperative SCZ	**56** / 44	**48** / 52
Subsample 6 (S6), Uncooperative PTSD	**52** / 48	**38** / 62
Subsample 7 (S7), Uncooperative DEP	**55** / 45	**49** / 51

Note. Correct classifications are highlighted by bold format.

ON RANDOM RESPONDING AND FEIGNING

Table 6. Additional Analyses Examining the Responses Provided by Human Random Responders: Non-Random Patterns.

	S2	S3	S4	S5	S6	S7	Combined (S2 to S7)
Highest Chi^2 value	57.9	57.0	35.8	33.7	38.6	29.5	145.9
Item #	1	1	19	28	16	25	1
% of *True*	68%	68%	53%	55%	29%	57%	55%
% of *False*	24%	23%	41%	37%	60%	30%	30%
% of *Doesn't Make Sense*	8%	9%	6%	8%	10%	13%	15%

Notes. All Chi^2 values reported in this table are statistically significant at $p < .001$. The highest Chi^2 value refers to the highest Chi^2 value found across all 27 closed-ended items under investigation. For each subsample, the frequency of endorsement of each response option for the item with the Chi^2 value is reported. S1 (Subsample 1, $n = 100$) = Computer-Generated Data; S2 (Subsample 2, $n = 100$) = Blind Random Responders; S3 (Subsample 3, $n = 100$) = Real-Life Random Responder; S4 (Subsample 4, $n = 100$) = Uncooperative NP; S5 (Subsample 5, $n = 100$) = Uncooperative SCZ; S6 (Subsample 6, $n = 100$) = Uncooperative PTSD; S7 (Subsample 7, $n = 100$) = Uncooperative DEP.

ON RANDOM RESPONDING AND FEIGNING

Figure 1. Graphical Representation (Box Plot) of Sim-RRS Scores across Study 2 Subsamples.
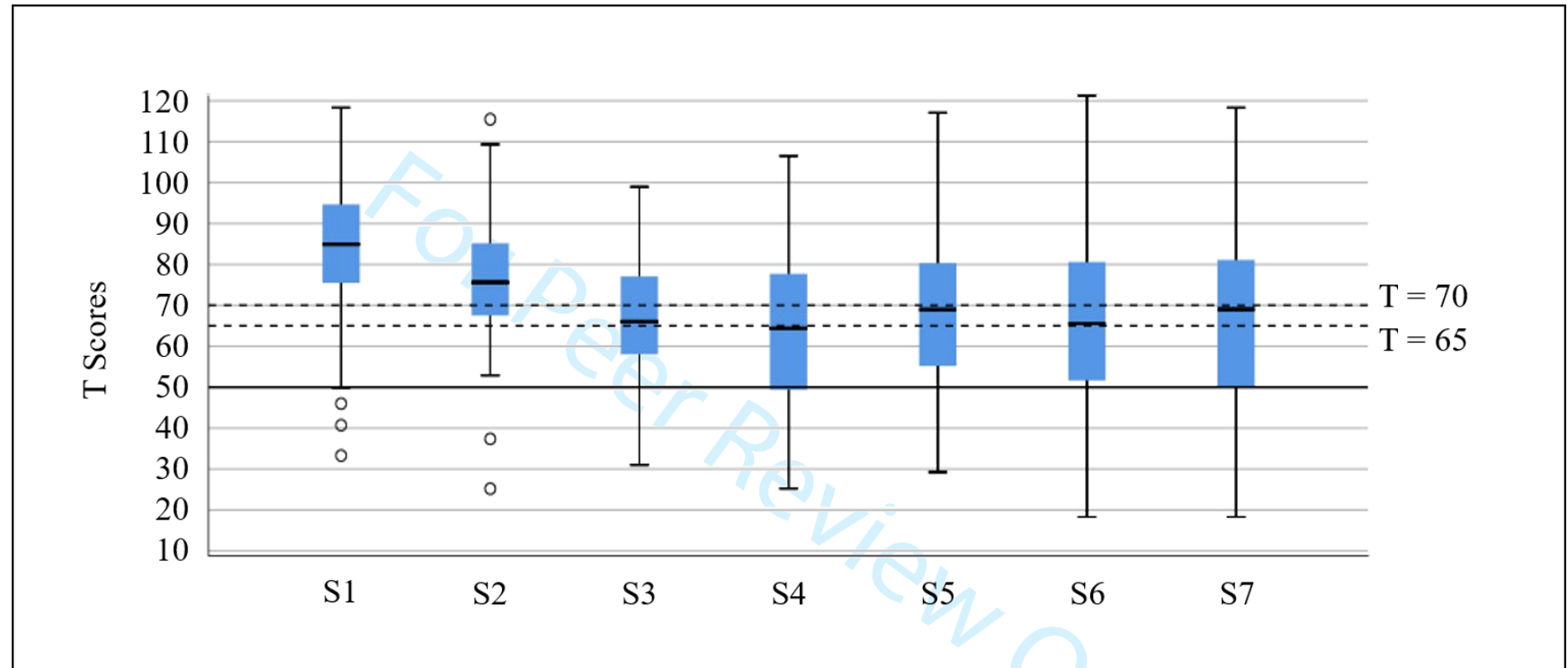
Notes. $N = 700$. S1 (Subsample 1, $n = 100$) = Computer-Generated Data; S2 (Subsample 2, $n = 100$) = Blind Random Responders; S3 (Subsample 3, $n = 100$) = Real-Life Random Responder; S4 (Subsample 4, $n = 100$) = Uncooperative NP; S5 (Subsample 5, $n = 100$) = Uncooperative SCZ; S6 (Subsample 6, $n = 100$) = Uncooperative PTSD; S7 (Subsample 7, $n = 100$) = Uncooperative DEP.