Personality and Individual Differences 149 (2019) 1–7

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Exploring a proactive measure of making items of a personality questionnaire resistant to faking: An employee selection setting

Wahyu Widhiarso[a,b,*], Rolf Steyer[a], Hamdollah Ravand[c]

[a] Friedrich Schiller University Jena, Germany
[b] Universitas Gadjah Mada, Indonesia
[c] Vali-e-Asr University of Rafsanjan, Iran

ARTICLE INFO

Keywords:
Employee selection
Faking
Latent variable modeling
Personality measure

ABSTRACT

This study examined resistance of item parameters of a personality questionnaire to faking. We proposed seven construct-irrelevant item attributes (CIIAs), each composing of two opposite values, representing degrees of adherence to the respective CIIAs (e.g., based on the time frame embedded in the stem, items were divided into past and present). We developed a personality inventory that measured conscientiousness. Items of the questionnaire were written to embody all the seven CIIAs. Participants of the study (N = 543) were divided into two groups: honest and faking groups. Those in the honest group were instructed to respond honestly whereas those in the faking groups were asked to respond as if they were job applicants. Multi-group latent variable modeling was used to examine the score differences of the participants' responses to the items of the conscientiousness section of the NEO FFI. It was found that items on two of the CIIAs were resistant to faking. Applying a Wald test of equality of item parameters, we found that the items did not show differential item functioning in either of the attributes.

1. Introduction

The answers to one of various questions about faking on personality scales have reached a consensus: people can fake their responses to the items of personality inventories (Viswesvaran & Ones, 1999). Therefore, research on faking behavior should shift from a *retroactive* to a *proactive* paradigm (Fan et al., 2012). In the retroactive paradigm, faking is allowed to occur during the test administration process. There are certain techniques (e.g., faking correction) which can be employed to refine the obtained scores to exclude the effect of faking. In contrast, in the proactive paradigm, researchers develop instruments in certain ways so as to make them more resistant to faking. There have been a few attempts to study the effect of different factors on making test items more faking-proof including modifying test format (Jackson, Wroblewski, & Ashton, 2000) or response format (Kubinger, 2009), composing items with similar content placed randomly (e.g., McFarland, Ryan, & Ellis, 2002), warning the participants (e.g., Dwight & Donovan, 2003), and using implicit measures (Vecchione, Dentale, Alessandri, & Barbaranelli, 2014). To a much lesser extent, the proactive impact of writing item has been investigated (e.g., Vasilopoulos & Cucina, 2006). In all the manifestations of the proactive approach, scale developers take preemptive measures to deal with faking.

The present study is an attempt at exploring the effect of the proactive approach via identifying item characteristics of a personality inventory that make them less vulnerable to faking. Developing faking-impervious instruments should focus on item level (e.g., by modifying the item stems) rather than test level (e.g., by developing test of specific response formats) features. The rationale behind the item- rather than test-level preemptive strategies is that people fake their responses to items, not to the entire scale. In other words, individuals who fake, usually employ a specific strategy which impacts their responses to only a subset of items. Even individuals who resort to maximum faking strategy do not necessarily obtain the maximum score because not all items are faked (Ziegler, 2011). Consequently, examining individuals' responses at the level of item rather than test is preferable because this approach is more informative in conceptualizing the faking process (Goffin & Boyd, 2009).

Besides content, items of personality inventories may have various other features. Each item may have a unique attribute that distinguishes it from the other items even from those measuring similar traits. Some items are more attractive than others thus they provide individuals' with opportunities to exaggerate positive impression. Among feature according to which items can be categorized are their context (e.g., at work vs. at school), direction (e.g., positively vs. negatively worded)

* Corresponding author at: Universitas Gadjah Mada, Jl. Humaniora No. 1 Bulaksumur, Yogyakarta, Indonesia.
E-mail address: wahyu_psy@ugm.ac.id (W. Widhiarso).

https://doi.org/10.1016/j.paid.2019.05.040
Received 18 August 2018; Received in revised form 21 May 2019; Accepted 22 May 2019
Available online 30 May 2019
0191-8869/ © 2019 Elsevier Ltd. All rights reserved.

and time framing (e.g., past vs. future oriented).

Faking can be influenced by the nature of items (Kuroyama, Wright, Manson, & Sablynski, 2010), including item content and features. Individuals may not be inclined to fake their responses to particular items because these items do not induce them to fake for some reasons. Job applicants, for example, might be reluctant to fake when they respond to items that are not related to the job offered or when their responses to the items are easily verifiable. Faking on items of personality scales is affected by both item content and item features (attributes). To make items of a personality inventory resistant to faking, one need not modify item content since it may alter the construct being measured. Instead, researchers can modify the item attribute without altering the item content.

1.1. Content irrelevant item attributes

In the present paper item attributes are referred to as *construct-irrelevant item attributes* (CIIAs) because they do not pertain to the actual content intended to be measured by the item, rather they add some elements to make the item more specific. This notion is similar to what is called *item contextualization* (Schmit, Ryan, Stierwalt, & Powell, 1995) or *item attribute* (Graham, McDaniel, Douglas, & Snell, 2002; Mael, 1991) in the literature. Each CIIA usually consists of two opposite categories describing the degree to which items adhere to the respective attribute called *attribute value*. Various types of CIIAs have been explored (e.g., Asher, 1972; Mael, 1991). Some researchers (e.g., Graham et al., 2002; Vasilopoulos & Cucina, 2006) have examined whether certain types of CIIAs can make the items resistant to faking (e.g., Graham et al., 2002; Vasilopoulos & Cucina, 2006). However, prior efforts involved a limited and very specific type of item stems such as *extremity* (Nye, Newman, & Joseph, 2010), *descriptiveness* (Bäckström, Björklund, & Larsson, 2009), *directions*, (Sliter & Zickar, 2014) and *obviousness* (Herkov, Gordon, Gynther, & Greer, 1994).

The current study employs seven CIIAs: *verifiability, source of information, time frame, discreteness, definiteness, internality,* and *context specificity*. Most of the CIIAs used in this study were adopted from Mael's (1991) taxonomy of item attributes for biodata, except for *discreteness* and *context specificity* that were adopted from Graham et al. (2002) and Schmit et al. (1995), respectively. Owing to the fact that the current study uses the personality inventory, the definition of each CIIAs is modified to make them match the type of instrument being used (i.e., Likert format). For example, in its original concept, *definiteness* was defined to represent whether or not the range of response options follows an obvious continuum (Doll, 1971). For the present study, this definition was modified as whether item stems used word modifier that reflected an explicit degree regarding to trait being measured. Hence, the focus of *definiteness* attribute is changed from continuity of response options (or anchors) to continuity of item stems. Items such as “I frequently…” or “I am a little...” were included in the *indefinite* category, since both items use word modifiers that express an explicit degree (i.e., frequently, little).

Each CIIA is divided into two attribute values, representing opposite characteristics. This division is made to facilitate testing of the resistance of each attribute value to faking. This is because the current study hypothesizes that some attribute values are more resistant to faking than their corresponding opposite values. Items can be considered resistant to faking if they exhibit equal item parameters across the honest and faking groups. Differential Item Functioning (DIF) occurs when individuals in different subgroups with equal standing on the latent construct being measured have different probabilities of endorsing any given item. Items are flagged for DIF if they have different response functions for different groups because individuals from different groups have different probabilities of endorsing a particular option although they have identical levels of the trait (Camilli & Shepard, 1994). Individuals who intend to fake are more likely to endorse items that are susceptible to faking than individuals who do not have such intention, even though they stand on equal trait levels. Hence, DIF analysis could be helpful in determining the effect of faking on the functioning of items as well as examining whether items of the scale are resistant to faking.

1.2. The purpose of the current study

Previous works have shown that resistance of the personality inventory items to faking depends on how the items have been written (Becker & Colquitt, 1992; Kuroyama et al., 2010). However, there is no consensus as to what makes certain item stems more resistant to faking than others. Thus, the purpose of the current study was to explore resistance of items to faking using dichotomous item features: CIIAs. It is hypothesized that items representing one attribute value of the proposed CIIAs are more resistant to faking than their corresponding pairs. Specifically, it was hypothesized that the *more verifiable*, *second-hand*, *past*, *non-discrete*, *definite*, *externality*, and *context-unspecified* items were more resistant to faking than the less *verifiable*, *first-hand*, *future-hypothetical*, *discrete*, *indefinite*, *internality*, and *context-specified* items. For the purpose of the present study, items were taken as resistant to faking if they were not flagged for DIF.

2. Method

2.1. Participants

Participants of the present study were 543 students (412 females, 131 males) who were taking one of the several undergraduate courses in psychology. Participants were selected using non-random sampling. The mean participants' age was 21 years (SD = 2.07) with left skewed distribution.

2.2. Instruments

2.2.1. The Original Version of the Conscientiousness Inventory (OCI)

Conscientiousness Inventory is part of The NEO Five-Factor Inventory (NEO-FFI) developed by Costa and McCrae (1989). It is used to measure individuals' level of consciousness. This OCI is contains 12 Likert items on a 5-point scale with response options ranging from “strongly disagree” to “strongly agree”. The scale demonstrated acceptable reliability ($\alpha = 0.890$).

2.2.2. The Modified Version of Conscientiousness Inventory (MCI)

The MCI was developed by the authors to measure individuals' conscientiousness. This instrument consisted of 64 Likert type items on a 5-point scale with response options ranging from “strongly disagree” to “strongly agree”. The items of this instrument were written with certain restrictions so that the item stems embodied the proposed CIIAs. The items on this scale represent seven item attributes: *verifiability*, *source of Information*, *time frame*, *discreteness*, *definiteness*, *action*, and *context*. Examples of the items in this instrument are presented in Appendix. The instrument was piloted on a sample ($N = 173$) before being used in this study. As a criterion for inclusion, items which had higher correlations with the OCI part of the NEO-FFI. The internal consistency of the MCI is high ($\alpha = 0.97$). The most likely reason for this very high alpha value is the redundancy of items' content (Pesudovs, Burr, Harley, & Elliott, 2007). The dimensional analysis yielded an acceptable fit of unidimensional model ($\chi 2 = 2858.874$, df = 1952; $p < .001$; CFI = 0.959, TLI = 0.958, RMSEA = 0.042, WRMR = 1.130). Low RMSEA (below 0.06) and higher TLI and CFI (above 0.95) values generally reflect good model fit (Hu & Bentler, 1999).

2.3. Procedures

Participants were randomly assigned to the control (honest group)

and manipulated group (faking group). First and foremost, the participants in both groups received the OCI. In the next step, participants in the control group completed the MCI under the standard instruction; they completed the scale as honestly as possible. On the contrary, participants in the faking group completed the scale under modified instruction. They were asked to make the good impression and respond as if they were taking part in job recruitment.

2.4. Data analysis

The current study employed multi-group confirmatory factor analysis (MG-CFA) to detect items infected by DIF. MG-CFA involved the two groups: the honest group and the faking group. Since the main purpose of the study was to examine resistance of items to faking, we were interested in identifying DIF with respect to the item location parameters (i.e., uniform DIF). The use of MG-CFA to detect DIF items in the present study is different from the previous studies (e.g., Church et al., 2011).

In the previous studies, researchers usually compared model fit of the more and less restricted models. Conventionally, the more restricted model imposes equality constraints on the parameters of the items of interest while in the less restricted model parameters are free to vary across the groups. However, the present study tested the item location parameter differences between the two groups. The differences were obtained by subtracting item locations of the honest and the faking groups.

In the first phase of the analysis, a unidimensional model was fitted to the data of both groups, separately, by treating individuals' responses to each item as ordered categorical variables. Mplus 7.0 (Muthén & Muthén, 1998–2011) was employed to compute the item parameters of the model using WLSMV estimation approach that provides a single factor loading and several threshold parameters for each item. Goodness of fit statistics such as Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and Root Mean Square Error of Approximation (RMSEA) were used to determine model fit.

Literature shows that there is a close relationship between CFA and IRT when applied to categorical variables (Mislevy, 1986; Muthén & Asparouhov, 2001). Many studies provided best practices in conducting MG-CFA to detect the presence of DIF to gather information about the comparability or invariance of item parameter across groups in ordinal or Likert format data (e.g., Pendergast, von der Embse, Kilgus, & Eklund, 2017). Researchers (Kim & Yoon, 2011) have shown that MG-CFA is superior to the more traditional techniques such as Mantel-Haenszel test in terms of precision and sensitivity to detect the presence of DIF.

3. Results

Because a one-dimensional model is fit to the data, the measurement yields only a single factor score which represents the individuals' standing with respect to the construct being measured. Therefore, descriptive statistics described in this section rely only on raw scores in order to identify the resistant items under similar attribute values. The modified instruction successfully made the participants in the faking group elevate their scores. All mean sub-scores of the faking group are higher than those of the honest group.

3.1. Testing the model fit

Conducting MG-CFA in Mplus requires equivalence of the number of response options that have been chosen by individuals across the compared groups. The analysis is terminated when, for example, several participants in the faking group chose the fifth response option but no one in the honest group chose the same option. To address this problem, individuals' responses to the missed categories were collapsed into their nearest categories. There is no standard rule for collapsing response categories, but the common practice is to merge categories with zero response rates with their closest category (Table 1).

Table 1
Item location parameter for each item attribute.

Item attributes	Honest group		Faking group		Wald statistic	N items
	M	SD	M	SD		
Verifiability						
More verifiable	1.34	0.21	1.95	0.97	36.582 (5)*	5/3
Less verifiable	1.10	0.33	0.76	0.78	98.239 (5)*	5/4
Source of inform.						
1st hand	1.11	0.30	0.71	0.22	50.208 (5)*	5/2
2nd hand	1.23	0.14	1.11	0.20	3.582 (5)	5/0
Time frame						
Past	1.22	0.30	1.25	0.28	9.610 (5)	5/1
Future	0.86	0.11	0.13	0.84	144.016 (5)*	5/4
Discreteness						
Non-discrete	1.08	0.12	0.75	0.27	26.828 (4)*	4/2
Discrete	1.35	0.18	1.45	0.71	29.387 (4)*	4/3
Definiteness						
Definite	1.15	0.14	0.91	0.23	19.292 (4)*	4/2
Indefinite	1.19	0.26	0.63	0.20	29.539 (4)*	4/4
Internality						
External	1.31	0.30	1.16	0.46	14.763 (4)*	4/1
Internal	1.10	0.43	0.00	0.36	246.325 (4)*	4/4
Context specificity						
Specific	1.17	0.17	0.17	0.29	151.592 (5)*	5/5
Non-specific	1.20	0.12	1.54	0.46	17.501 (5)*	5/2

Note. N items represents number of items/number of items affected by DIF.
* $p < .001$.

The results of the CFA for each group indicate that the baseline model had an acceptable fit in both the honest group ($\chi^2 = 2858.874$, $df = 2773$; $p < .001$; CFI = 0.967, TLI = 0.966, RMSEA = 0.033) and faking group ($\chi^2 = 2840.501$, $df = 2773$; $p > .05$; CFI = 0.997, TLI = 0.997, RMSEA = 0.010). Therefore, the baseline model can be perceived as a relatively stable measurement model that can be used for subsequent analysis. The results of the MG-CFA also yielded acceptable model fit ($\chi^2 = 6544.510$, $df = 5594$; $p < .001$; CFI = 0.981, TLI = 0.981, RMSEA = 0.010).

3.2. Examining the presence of differential item functioning (DIF)

As described in the Analysis section, to tests for DIF, the MG-CFA was repeated fourteen times, each with a different omnibus Wald test that examined whether sets of items within CIIAs exhibited different item location parameters. Only two (i.e., *second-hand* and *past items*) of the fourteen Wald test results exhibited non-significant results, indicating that the location parameter of the items for those attribute values are equal across the two groups. A closer examination of the item location differences indicated that 27 of the 64 items exhibited DIF. Several items that represented similar attribute values were displaying DIF, except for the items in the *second-hand* attribute value.

The results showed that only *second-hand* and *past* items are resistant to faking. Items in five attribute values (e.g., *more verifiable*, *discrete*) performed below the expectation in showing resistance to faking. However, the number of items displaying DIF is in line with what we expected. For example, the number of *more verifiable* items that exhibited DIF is lower than *less verifiable* items. Overall, *more verifiable*, *discrete*, *definite*, *externality*, and *context-unspecified* items were relatively more resistant to faking than their opposite counterparts (i.e., *verifiable*, *discrete*, *indefinite*, *internality*, and *context-specified*, respectively).

4. Discussions

This study compared resistance of attribute values of CIIAs to faking. The results showed that *past* and *second-hand* items are resistant to faking, suggesting that personality inventory items that figure past

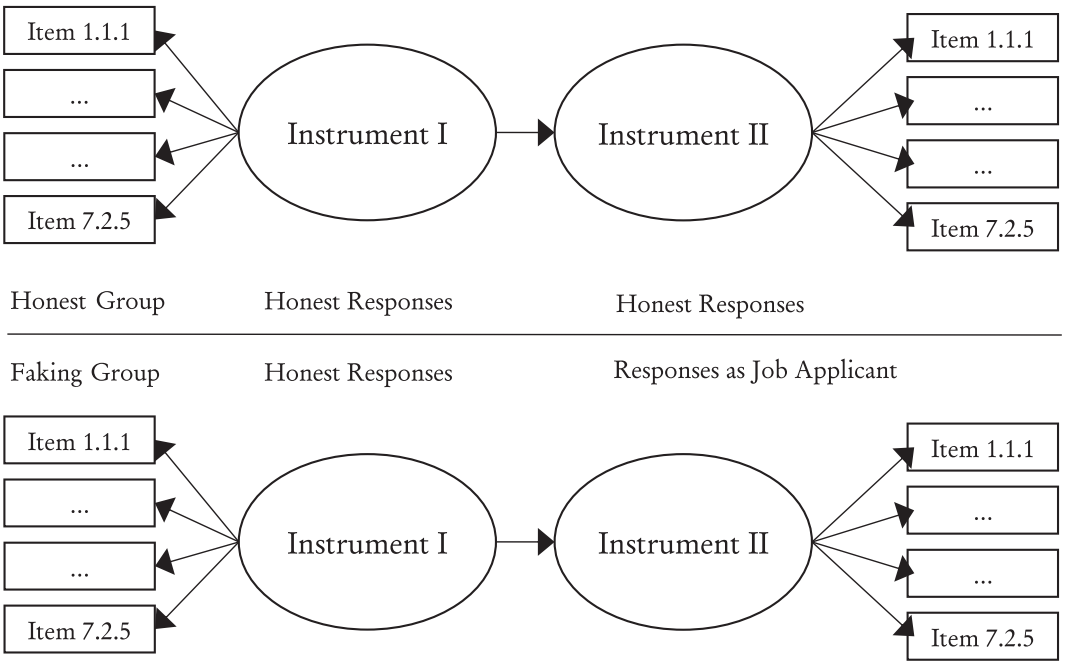

Fig. 1. Multigroup CFA with covariate.

orientation and others' views are resistant to faking. However, items that featured the rest of the attribute values showed susceptibility to faking (Fig. 1).

The results support the claim that past behavior questions are more resistant to faking than situational questions (Levashina & Campion, 2007). *Past behavior* items ask individuals to portray a historical and more real situation than *future* items do, which rely on behavior on imaginative or even nonexistent situations. Besides, responses to those past items which are more verifiable than *future* items might prevent individuals from faking. This is mostly true about those who think that there is another stage in the job selection process (e.g., interview) that affected their responses. *Future* items assess individual intention or willingness to perform. Because future items rely on an intention, individuals can fully manage their responses to these items. Individuals could express wrong intention to perform productive behavior as indicated by all *future* items on the instrument used in this study (e.g., "I try to do my work carefully").

The other attribute value that showed resistance to faking was the *second-hand*. The results of this study supports the previous studies (Becker & Colquitt, 1992), which found that among the CIIAs examined, only *second-hand* items showed resistance to faking. Considering the verifiability, *second-hand* items are more verifiable than *first-hand* items, therefore, responses to these items are easier to corroborate. Accordingly, items with this feature are more resistant than items with the corresponding opposite values. However, the findings of this study are in contrast with Graham's et al. (2002) finding in that he found both *first-* and *second-hand* items were equally susceptible to faking as indicated by the equivalence of the validity of the items that featured *first-* and *second-hand* attributes when faking occurred. The finding of this study challenges the arguments that *second-hand* items are more susceptible to faking than *first-hand* items on the grounds that *second-hand* items add speculative subjectivity, making the items less objective and hence more easily faked (see Vasilopoulos & Cucina, 2006).

Besides the two attribute values described above, all the attribute values were more or less susceptible to faking. However, some attribute values were relatively more resistant to faking than their corresponding opposite values. This finding is consistent with the previous studies (e.g., Becker & Colquitt, 1992; Graham et al., 2002; Harold, McFarland, & Weekley, 2006) which found that more *verifiable* items are more relatively resistant to faking than less *verifiable* ones, although the performance of both attribute values to preventing faking was not high. The previous studies used biographical measures that gathered background or personal experience data, whereas the present study used a personality inventory that measured a personality trait.

Another finding of the present study that *definite* items were more resistant to faking than *indefinite* items is also consistent with the extant literature (Graham et al., 2002). *Definite* items provide a clear and specific question that makes individuals cautious when responding to them. On the contrary, *indefinite* items usually involve continuous modifiers (e.g., to some extent, almost) which provide a grey situation that makes the items less evaluative. *Indefinite* items make the degree of assessed indicator become blurred but offer a clear path to faking. This argument also applies to the finding that *discrete* items are relatively more resistant than *non-discrete* items.

It was also found that *external* items are more relatively resistant to faking than *internal* items because they concern explicit events, as opposed to *internal* items which concern events inside the individual (e.g., thought). This finding supports findings of the previous studies (e.g., Becker & Colquitt, 1992). Internal activities such as expecting, choosing, desiring, and intending to behave (Bean & Eaton, 2000) are subjective. Everyone will be able to defend their arguments, as long as they convey positive or desirable behaviors. The susceptibility of *internal* items to faking is caused by the fact that individuals can defend their arguments when they fake their responses to these items.

It was also found that *non-specific* items are more resistant to faking than *specific* items. *Non-specific* items assessed indicators that were not clearly associated with the offered job whereas *specific* items represented transparency of their connection. Individuals may be more motivated to fake *specific* items because the implications of responses to such items are more apparent and direct than *non-specific* items (Becker & Colquitt, 1992). Examining both attribute values, especially using the personality inventory, might relatively new approach in faking research area because most of the prior studies merely contrasted obvious and subtle items (Brems & Harris, 1996; Wierzbicki, 1997).

The practical implication of this research is that use of appropriate item stems in the personality inventory can reduce faking behavior in personnel selection settings. The CIIAs proposed in this study provide a conceptually useful framework for generating items to measure various

personality traits. Although the theoretical basis of some CIIAs proposed in this study is taken from biographical measurement literature, this study demonstrated applicability of these features to the personality inventory. The Results of fitting unidimensional model to the data showed that although the Personality Inventory composed of items with various CIIAs, it still measured the single trait of interest. Hence, instrument developers can include one or several attribute values (i.e., CIIAs) into the item stems to make them more resistant to faking. CIIAs are generic classification so that they could be used in various personality constructs (e.g., extroversion) or personality inventory with different response formats (e.g., Thurstone format).

4.1. Study limitations

There are several limitations to this study that need to be addressed. The collinearity among CIIAs precluded testing the latent score differences among CIIA. Collinearity can produce overestimated parameters or inflate the standard errors that may make the estimated parameters become untrustworthy. Item content redundancy across CIIAs might have led to the presence of collinearity. Therefore, future studies should develop an instrument composed of various CIIAs that assess different content areas. As an alternative, further studies can use an instrument with similar composition using different analysis techniques. For example, modeling method effect to examine the effect of faking on item parameters.

Appendix 1. Scale development procedures

This study used a personality inventory to measure two facets of conscientiousness (*achievement* and *dependability*; Roberts, Chernyshenko, Stark, & Goldberg, 2005). The process of validation was performed using literature review, semantic and content validation by a panel of experts. Respondents were asked to indicate their degree of agreement to 64 statements (i.e., items). All items were scored on a 5-point Likert scale, ranging from 1 = "strongly agree" to 5 = "strongly disagree". This inventory represents seven CIIAs, each of which has two attribute values represented by four or five item pairs: The CIIAs of *verifiability*, *source of information*, *time frame,* and *context specificity* are represented by five item pairs each, while the CIIAs of *discreteness*, *continuity*, and *internality* are represented by four items pairs each (see Fig. 1).

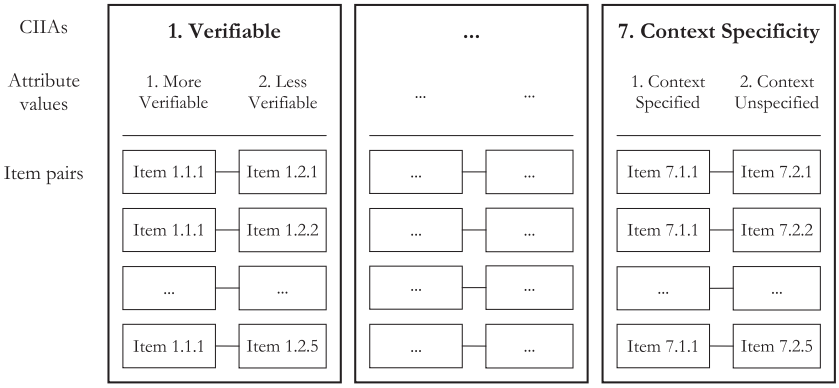

Fig. 1. Item specification of developed scale for current study.

The personality inventory used in this study was validated with a sample of university students (N = 173). Various kinds of item analyses were conducted to select items based on their psychometric properties. The criteria for including an item in the final instrument were factor loadings—which reflect how well an item can discriminate among individuals for the latent trait—and the correlation of the item to an external criterion. In this study, the conscientiousness subscale from the NEO Five Factor Inventory (NEOFFI) was used as an external criterion. Of the 80 items in the item pool, six items exhibited low discrimination (factor loading < 0.30) and eight items had low criterion validity (correlation < 0.30). Because in the current study item pairs represent attribute values, any item pair was also dropped if at least one of the two items did not fulfill the criteria for item inclusion mentioned above. Hence, a total of 16 items were excluded from the final scale. The internal consistency of the scale (composed of 64 items) was very high (Cronbach's alpha = 0.970). The most plausible reason for this high alpha can be the similarity of items' content in the inventory (Pesudovs et al., 2007), because similar content (e.g., the degree to which persons follow their plans and display achievement and commitment) is used in items with different CIIAs. Two models were proposed: a one-dimensional model in which all items load on the same factor and a multidimensional seven-factor model that separated items according to their attributes. The results of analysis indicate that the one-dimensional model yielded an acceptable fit than multidimensional model ($\chi 2 = 2858.874$, df = 1952; $p < .001$; CFI = 0.959, TLI = 0.958, RMSEA = 0.042, WRMR = 1.130). A lower RMSEA (below 0.06) and higher TLI and CFI (above 0.95) are generally accepted as reflecting good model fit (Hu & Bentler, 1999).

Appendix 2

Table 1
Classification of items by attribute, including sample items.

Item attribute	Description and sample item
Verifiability. Assesses whether responses can be corroborated from independent sources	
1. More verifiable	Responses to items in this value can be corroborated from independent sources; this includes sources that are difficult to access. Example: "I have numerous professional achievements."
2. Less verifiable	Responses to items in this value cannot be corroborated from independent sources, since they inquire about individuals' thoughts, feelings, or opinions. Example: "I think I achieved excellence in my work this year."
Source of Information. Asks who or what provided the information about the individual	

(*continued on next page*)

Table 1 (*continued*)

Item attribute	Description and sample item
1. First-hand source	An individual provides an evaluation or judgment about him or herself. Example: "I am a person who works diligently."
2. Second-hand source	Asks for another person's evaluation or judgment of the individual. Example: "My supervisor rates me as a diligent worker."
Time frame. Classifies item stems according to time frame	
1. Past–Present	Asks individuals to report on their own experiences that happened in the past or present. Example: "I complete my projects according to plan."
2. Future–Hypothetical	Concerns individuals' willingness, hope, or intention to behave a certain way, as well as responses to hypothetical situations. Example: "I try complete my projects according to plan."
Discreteness. Classifies whether item stems contain one or two modifiers describing traits or attributes	
1. Discrete	Item stems contain two modifiers. Example: "Cautious describes me better than careless."
2. Non-discrete	Item stem contains a single modifier. Example: "I am a disciplined worker."
Continuity. Classifies whether items represent attributes that can be placed along a clear continuum	
1. Indefinite	The degree of the reported attribute is obviously determined and involves modifiers such as "usually" or "to a great extent." Example: "Sometimes I do my job in a professional manner."
2. Definite	The behavior is clearly and specifically explained, but without modifiers. Example: "I do my job in a professional manner."
Internality. Assesses whether the item represents covert (i.e., internal) or overt (i.e., external) expression	
1. Internality	Items concern events inside the individual, and assess individual attitudes, opinions, and emotional reactions to events. Example: "Mood should not affect an individual's effort when finishing a job."
2. Externality	Items concern events outside the individual, such as behavior, past experiences, or explicit expression. Example: "I focus on my job even my mood is poor."
Context specificity. Classifies items with regard to whether or not the context is specified.	
1. Context-specified	Item stem refers to a specific context (e.g., specific location, time, situation) Example: "I do not to waste time when working."
2. Context-unspecified	Item stem is more general, with an unclear or unspecific context. Example: "I do not to waste my time."

Appendix 3. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.paid.2019.05.040.

References

Asher, J. J. (1972). The biographical item: Can it be improved? *Personnel Psychology, 25*(2), 251–269.

Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality, 43*(3), 335–344. https://doi.org/10.1016/j.jrp.2008.12.013.

Bean, J. P., & Eaton, S. (2000). A psychological model of college student retention. In J. M. Braxton (Ed.). *Rethinking the departure puzzle: New theory and research on college student retention*. Memphis: University of Vanberbilt Press.

Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form - An analysis along several dimensions of item type. *Personnel Psychology, 45*(2), 389–406.

Brems, C., & Harris, K. (1996). Faking the MMPI-2: Utility of the subtle-obvious scales. *Journal of Clinical Psychology, 52*(5), 525–533.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items.* Thousand Oaks, CA: Sage Inc.

Church, A. T., Alvarez, J. M., Mai, N. T. Q., French, B. F., Katigbak, M. S., & Ortiz, F. A. (2011). Are cross-cultural comparisons of personality profiles meaningful? Differential item and facet functioning in the Revised NEO Personality Inventory. *Journal of Personality and Social Psychology, 101*(5), 1068–1089. https://doi.org/10.1037/a0025290.

Costa, P. T., & McCrae, R. R. (1989). *The NEO-PI/NEO-FFI manual supplement.* Odessa, FL: Psychological Assessment Resources.

Doll, R. E. (1971). Item susceptibility to attempted faking as related to item characteristic and adopted fake set. *Journal of Psychology, 77*(1), 9–16.

Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*(1), 1–23.

Fan, J., Gao, D., Carroll, S. A., Lopez, F. J., Tian, T. S., & Meng, H. (2012). Testing the efficacy of a new procedure for reducing faking on personality tests within selection contexts. *Journal of Applied Psychology, 97*(4), 866–880. https://doi.org/10.1037/a0026655.

Goffin, R. D., & Boyd, A. C. (2009). Faking and personality assessment in personnel selection: Advancing models of faking. *Canadian Psychology, 50*(3), 151–160. https://doi.org/10.1037/A0015946.

Graham, K. E., McDaniel, M. A., Douglas, E. F., & Snell, A. F. (2002). Biodata validity decay and score inflation with faking: Do item attributes explain variance across items? *Journal of Business and Psychology, 16*(4), 573–592. https://doi.org/10.1023/A:1015454319119.

Harold, C. M., McFarland, L. A., & Weekley, J. A. (2006). The validity of verifiable and non-verifiable biodata items: An examination across applicants and incumbents. *International Journal of Selection and Assessment, 14*(4), 336–346.

Herkov, M. J., Gordon, R. A., Gynther, M. D., & Greer, R. A. (1994). Perceptions of Mmpi item subtlety - influences of age and ethnicity. *Journal of Personality Assessment, 62*(1), 9–16.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*. https://doi.org/10.1080/10705519909540118.

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance, 13*(4), 371–388.

Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(2), 212–228. https://doi.org/10.1080/10705511.2011.557337.

Kubinger, K. D. (2009). Three more attempts to prevent faking good in personality questionnaires. *Review of Psychology, 16*(2), 115–121.

Kuroyama, J., Wright, C. W., Manson, T. M., & Sablynski, C. J. (2010). The effect of warning against faking on noncognitive test outcomes: A field study of bus operator applicants. *Applied Human Resources Management Research, 12*(1), 59–74.

Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: Development and validation of an Interview Faking Behavior scale. *Journal of Applied Psychology, 92*(6), 1638–1656.

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology, 44*(4), 763–792.

McFarland, L. A., Ryan, A. M., & Ellis, A. (2002). Item placement on a personality measure: Effects on faking behavior and test measurement properties. *Journal of Personality Assessment, 78*(2), 348–369.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational and Behavioral Statistics, 11*(1), 3–31. https://doi.org/10.3102/10769986011001003.

Muthén, B., & Asparouhov, T. (2001). *Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus.* Los Ange les: University of California and Muthén & Muthén.

Muthén, B., & Muthén, L. K. (1998–2011). *Mplus (Version 7.0) [Computer software].* Los Angeles, CA: Muthén & Muthén.

Nye, C. D., Newman, D. A., & Joseph, D. L. (2010). Never say "always"? Extreme item

wording effects on scalar invariance and item response curves. *Organizational Research Methods, 13*(4), 806–830. https://doi.org/10.1177/1094428109349512.

Pendergast, L. L., von der Embse, N., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology, 60*, 65–82. https://doi.org/10.1016/j.jsp.2016.11.002.

Pesudovs, K., Burr, J. M., Harley, C., & Elliott, D. B. (2007). The development, assessment, and selection of questionnaires. *Optometry & Vision Science, 84*(8), 663–674. https://doi.org/10.1097/OPX.1090b1013e318141fe318175.

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology, 58*(1), 103–139. https://doi.org/10.1111/j.1744-6570.2005.00301.x.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80*, 607–620. https://doi.org/10.1037/0021-9010.80.5.607.

Sliter, K. A., & Zickar, M. J. (2014). An IRT examination of the psychometric functioning of negatively worded personality items. *Educational and Psychological Measurement, 74*(2), 214–226. https://doi.org/10.1177/0013164413504584.

Vasilopoulos, N. L., & Cucina, J. M. (2006). Faking on non-cognitive measures: The interaction of cognitive ability and test characteristics. In R. L. Griffith, & M. H. Peterson (Eds.). *A closer examination of applicant faking behavior*. Greenwich, CT: Information Age Publishing.

Vecchione, M., Dentale, F., Alessandri, G., & Barbaranelli, C. (2014). Fakability of Implicit and Explicit Measures of the Big Five: Research findings from organizational settings. *International Journal of Selection and Assessment, 22*(2), 211–218. https://doi.org/10.1111/ijsa.12070.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*(2), 197–210. https://doi.org/10.1177/00131649921969802.

Wierzbicki, M. (1997). Use of subtle and obvious scales to detect faking on the MCMI-II. *Journal of Clinical Psychology, 53*(5), 421–426. https://doi.org/10.1002/(sici)1097-4679(199708)53:5<421::aid-jclp2>3.0.co;2-q.

Ziegler, M. J. (2011). Applicant faking: A look into the black box. *The Industrial-Organizational Psychologist, 49*(1).