

## Item context effects on Big Five personality measures

VESNA MLINARIČ and ANJA PODLESEK

Item specificity and contextual cues are one of the potential sources of measurement error in personality measures. The term *frame of reference* represents an implicit or explicit context that affects individual's answers while answering personality items. The paper discusses the consequences of standardizing the contextual understanding of items on the psychometric characteristics of the Big Five personality measures. Although highly related, context-specific measures can result in different personality profiles compared to general measures. Furthermore, different types of contextualization of personality questionnaires can result in different personality profiles. Item contextualization may also affect the reliability of personality measures, even though the effect is not entirely clear and different studies report different effects. When the context is conceptually relevant to the criterion measure, adding item specificity may also affect predictive validity. Research therefore implies some potential benefits for personality assessment through using context specific measures. However, there are still some open questions, namely contextualised measures can also be susceptible to potential distortions, which are further described in this paper. At the end we discuss potential implications and limitations of context-related personality assessment.

*Key words:* personality measures, item context, factor structure, test reliability, test validity

The most widely used personality questionnaires are based on the trait paradigm, which implies relatively stable behavioural and emotional patterns. This assumption, however, has often been criticized and empirically called into question. For example, recent research has shown that trait questionnaires and aggregated momentary assessments of personality are highly related, but do not show convergence (Augustine & Larsen, 2012). Therefore evidence exists that general self-report personality questionnaires capture different aspects of one's personality, depending on various measurement circumstances. Previous research (e.g., Gordon & Holden, 1996, 1998) has also shown that individuals use different strategies for answering items in personality questionnaires. These strategies may be affected implicitly and/or explicitly by test instructions and questionnaire form. They affect individual personality trait scores and the measurement properties of the questionnaire. One of these influences can also be contextual or situational cues, i.e., the frame-of-reference effect.

The well-known trait vs. situation debate, which highlights the question whether stable factors or situational ones

influence behaviour, has a long history in psychology. One of the more influential theories – the theory of conditional dispositions (Wright & Mischel, 1987) – also serves as a theoretical background for constructing situation-based personality assessment. Wright and Mischel acknowledged the impact of traits as predictors of behaviour, but also emphasized certain limitations and potential measurement improvements. They stated that although traits can serve as a basis of behaviour, their manifestation depends upon the situational characteristics. Individuals may behave very consistently across a range of similar situations but can also behave in a very different way in different situations. For example, an individual may consistently display extraverted behaviours at home, but fail to display the same trait characteristics in work-related situations. Although there is variability in trait manifestation across different situations, consistency and predictability can arise in similar situations. Therefore the predictive value of personality traits may increase when situational factors are taken into account.

Another important factor in answering personality items is accessibility of the self-concept at the time of answering. For example, responding to items related to a work situation with one's "employee self-concept" could be of greater relevance and more easily accessible compared to other self-concepts, and it may also have a greater impact on the respondent's behaviour in work situations (Heller, Ferris, Brown, & Watson, 2009).

---

Vesna Mlinarič, University Rehabilitation Institute, Linhartova 5, 1000 Ljubljana, Republic of Slovenia. E-mail: za.vesnamlinaric@gmail.com (the address for correspondence);

Anja Podlessek, Department of Psychology, University of Ljubljana, Slovenia.

The term *frame of reference* represents an implicit or explicit context that affects individual's answers. The effect can be caused by the situation itself, given through standard test instructions or by altering personality-test items. It has been proposed that under generic instructions different respondents may use different frames of reference when responding to general questionnaire items (Schmit, Ryan, Stierwalt, & Powel, 1995). For example, in an employee selection process some job applicants may feel that it is appropriate to give information only about how they are likely to behave at work, whereas others may be inclined to respond according to their general self-schema. Some of the participants may give an answer to a certain item according to their previous work experience, while others may refer to a generalized social situation. The variability of responses of different participants may therefore be attributed not only to individual differences in the measured construct, but also to the differences in the participants' interpretation of the item. It could be said that subgroups of participants actually do not rate or answer the same item, which according to Schmit and his co-workers (1995) increases the measurement error. When items and the measurement situation are not narrowly defined, participants may self-contextualize items (Schmit et al., 1995).

As previously stated, when answering general (non-situational) questionnaires, participants may vary in the frames of reference they use. This variability can appear at the between-subjects level (each participant uses a different frame of reference) and at the within-subjects level (one participant may shift between different frames of reference in different items measuring the same construct). Therefore, adding a specific frame of reference (e.g., work, school) should standardize item content interpretation across participants and should, at the individual level, support the use of a single frame of reference and increase the consistency of the participant.

*The effect of frame of reference on the values of personality measures*

In their initial study, Schmit and his co-workers (1995) compared general personality questionnaire (NEO-FFI, NEO-PI-R) scores to more situational specific personality measures by manipulating several administration conditions (applicant vs. general instructions) and item specificity (contextualized vs. general/non-contextual items). The authors assumed that responses of the participants in a general, situation non-specific assessment will be different from the ones in the selection context. Due to specific contextual cues in the selection context, respondents may be motivated to give socially desirable answers and may have specific frames of references in mind when describing their personality. The authors achieved higher item specificity by modifying items, i.e., by giving them a specific uniform frame of reference – at work or at school. For example, the item “I

try to be courteous to everyone I meet” was modified to “I try to be courteous to everyone I meet at work” (Schmit et al., 1995). They assumed that most individuals recognize to some extent that their behaviour at work or at school may be more constrained than their behaviour in some other general life situations, therefore the answers would differ with item type. Respondents (students) rated their agreeableness, conscientiousness and emotional stability consistently higher with work-specific personality measures than with general personality measures of the NEO-FFI. The only dimension with no evident differences was extraversion. Schmit and his co-workers (1995) proposed that such a result is a consequence of higher generalizability of extraversion across situations. Openness was not included in the study. At the facet level of the Conscientiousness scale the trend was similar – in the school-specific item condition respondents scored higher than in the non-contextual conditions.

Heller and his co-workers (2009) also found systematic differences in personality measures when comparing work and home personality. In comparison to their home personality, the participants reported to be significantly more conscientious and open to experience and less extraverted when answering according to their work context. Lievens, De Corte, and Schollaert (2008) similarly compared IPIP-50 personality dimension scores obtained in the general, work-specific, and school-specific context, and reported small effect sizes for item contextualization on all five dimensions and also on two facets of conscientiousness (on the achievement striving facet and the self-discipline facet). On all measures except agreeableness, where respondents had the highest average score in the generic condition, the respondents rated their work-related personality traits most positive.

In our own study (Mlinarič, 2012) we compared students' responses to a general and a school-specific version of the MINI IPIP questionnaire. We found that agreeableness and openness were rated higher and conscientiousness and neuroticism were rated lower in the general-personality version compared to the school-specific version of the items. To study the dimension of conscientiousness in more detail, we additionally used 60 items from the IPIP-NEO 300 questionnaire. Comparing the general and school-specific versions of these items, we found that the scores on the six conscientiousness facets differed for the two versions of items: competence, achievement striving, and self-discipline were rated higher when participants were responding to general personality items, and order, dutifulness, and deliberation were rated higher when participants were responding to school-specific items.

*The effect of frame of reference on reliability of personality measures*

The mentioned studies show that personality measures are sensitive to item specificity, which can affect the con-

ceptual integrity of the personality scales. The question arises whether in the context-specific versions the Big Five factors of personality are still being measured. Schmit and his co-workers (1995) tried to answer this question. They conducted confirmatory factor analyses for the one-dimensional and four-dimensional models of personality (openness was excluded from their analyses) and assessed model fit. The four-factor model of the NEO-FFI provided a better fit to the data than the one-factor model. The error variances for the Big Five model and the facet level of conscientiousness were significantly larger in the non-contextual than in the contextual condition, showing that item contextualization can give rise to an even clearer factorial structure. Robie, Schmit, Ryan, and Zickar (2000) examined the frame-of-reference effects on the NEO-PI-R Conscientiousness scale in a field study on applicants for police officer positions. In the frame-of-reference condition, where respondents answered items with at-work content, five of the six conscientiousness facets had significantly lower facet-level error variance. At the item level, three of the six facets had consistently larger error variances on the non-contextualized items. For each conscientiousness facet the work-specific items also had higher means.

Schmit and his co-workers (1995) assumed that differences in reliability between the frame-of-reference and general measures underlie the differences in the measurement error. According to their view, item contextualization leads to a higher between-person consistency and to a reduction in the measurement error. Lievens and his co-workers (2008) further explored the relation between context-specific items and reliability. They assumed that it is not only the reduced between-person variability that affects reliability, but that the reduced within-person variability is the main reason for the changes in reliability of the frame-of-reference measures. Namely, unstandardized situational content is linked to two sources of variability: between-person variability and within-person inconsistency. When the studies examine the frame-of-reference effect with the between-groups design (each group is presented with a different frame-of-reference condition), differences in reliability between groups reflect the frame-of-reference effects adequately only when all participants in one group refer to the same context. So, when studying context effects, it is important to assess not only to what level all the participants use the same context, but also how consistent each participant is when using one context. The within-person inconsistency also influences the reliability estimate.

Research with the within-subject design (repeated measures) has made it possible to assess the impact of the frame-of-reference effect on the level of an individual. Lievens and his co-workers (2008) used such a design and asked participants to answer items of the Conscientiousness scale (more specifically, the Achievement Striving and Self-Discipline scale) within both at-work and at-school context. Their findings had two crucial implications – Cronbach's alphas were

significantly higher for the scales rated with the at-school frame of reference than for the same scales rated with the at-work frame of reference. The reason for lower inconsistency of the scales in the at-work context may be the fact that participants were students who may be less familiar with the at-work context. Further on Lievens and his co-workers examined what effect the degree of within-person inconsistency would have on the reliability and validity of the studied personality measures. They sampled randomly without replacement from the responses of the total sample and used structural equation modelling to assess the psychometric characteristics. Reliability was the highest for the simulated situation where respondents would interpret a large number of items with the same frame of reference and it was lowest for the situation where respondents would switch their frame of reference for many of the items (i.e., for half of the items they would use a certain frame of reference and for the other half they would use a different one). The conclusion was that reliability is not affected as long as individuals are consistent within themselves or as long as the frames of reference between which the participants are switching have comparable reliabilities.

Further studies have shown different reliability patterns for the frame-of-reference measures. Bing, Whanher, Davidson, and VanHook (2004) and Bowling and Burns (2010) found no differences in the internal consistency of the general and work-specific measures. Reddock, Biderman, and Nguyen (2011) found the reliabilities of the Extraversion and Agreeableness scores to be significantly higher in the school-specific condition compared to the generic condition. In the study of Mlinarič (2012), several conscientiousness facets (competence, dutifulness, achievement striving, and self-discipline) showed higher internal consistency in the at-school condition than in the general condition. However, in the same study the general measures of extraversion and openness from the MINI IPIP had higher internal consistency than school-specific measures.

#### *The effect of frame of reference on predictive validity of personality measures*

As personality measures are often used as performance predictors in selection process, their low predictive value can be especially concerning. Therefore there is certainly a need for ways to improve the validity of such measures. Schmit and his co-workers (1995) assumed that providing the same frame of reference to all applicants (e.g., using items that specifically refer to behaviour at work) will improve the validity of personality measures for predicting performance measures. For example, respondents whose answers to personality items refer to their work experience may be providing information that is a better indicator of job performance, and consequently the predictive validity of personality measures can be improved.

According to previous research conscientiousness is most consistently related to performance measures (e.g., Chamorro-Premuzic & Furnham, 2003; De Fruyt & Merielde, 1996; Nettle & Robins, 2007), and therefore the majority of the frame-of-reference research has focused mostly on conscientiousness. In the second part of their study, Schmit and his co-workers (1995) examined the frame-of-reference effects on the facet level of the Conscientiousness scale. Items were modified by adding an “at school” tag. The school-specific conscientiousness measure had higher predictive value for grade point average (GPA), both when the questionnaire was answered with general instructions and when the respondents were instructed to answer as being in a job selection process. Schmit and his co-workers concluded that altering the items gave all respondents a common frame of reference, which has reduced error variance and increased validity. Bing and his co-workers (2004) replicated these findings. In their study, the frame-of-reference measures of conscientiousness had significantly higher criterion validity than the general measure of conscientiousness in predicting GPA.

Bing and his co-workers (2004) argued that cognitive ability could influence responses. When motivated to do so (e.g., in a job selection process), participants may tend to detect the wanted answer. Those with higher cognitive ability may find it easier to adjust their response according to the situation demands. But even when controlling for the cognitive abilities, the effect of contextualization led to a significant increase in validity for predicting GPA. Similarly they found that the effect of contextualization on validity remained significant even when facet reliability was specified as a covariate in a general linear model. Reddock et al. (2011) obtained similar results – in predicting GPA, validity was larger when conscientiousness items had a school-specific content than in the generic condition. In the study of Mlinarič (2012), the prediction of GPA was better with the school-specific openness measured with the MINI IPIP as a predictor than with the general openness measure.

Hunthausen, Truxillo, Bauer, and Hammer (2003) examined the frame-of-reference effects on validity on entry-level customer service managers in US Airports. Participants answered the NEO-FFI either in its original form or in a form where the reference to work before and during the inventory administration was made. They found that conscientiousness, extraversion and openness to experience were significantly correlated with job performance in the frame-of-reference condition but not in the standard condition. The frame-of-reference effect on validity was still supported after controlling for cognitive ability.

Lievens and his co-workers (2008) examined whether the frame of reference can improve predictive validity by itself or is it relevant which frame of reference is used. They found that the predictive value of conscientiousness for predicting GPA was different when participants answered the

questionnaire according to their work personality, school personality, or personality in general. The highest predictive value was obtained for the at-school condition. The use of a standard frame of reference had beneficial effects for the measurement properties only when contextualization matched the criterion, i.e., when the items were answered for the school context and GPA was used as a criterion. Also on the facet level, contextualization has shown beneficial effects on validity for conceptually relevant facets (e.g., conscientiousness, self-discipline and achievement striving; Lievens et al., 2008), which have been previously found to be related to the criteria of academic achievement (see e.g., Chamorro-Premuzic & Furnham, 2003), but not for broad personality factors measured with the IPIP 50. Lievens and his co-workers (2008) concluded that simply imposing a frame of reference is not enough. It is not beneficial to add item specificity by itself. Instead, equal importance should be placed on choosing the relevant frame – a frame of reference that is conceptually relevant to the criterion. This implies what consequences self-contextualization could bring if participants chose an incorrect frame of reference that did not conceptually overlap with the criterion – the predictive validity could be reduced. Lievens and his co-workers (2008) found that the increase in the validity of the context-specific version as compared to the general version was as much as .33. Most other studies, however, reported increases up to .16.

Although most research focused on the frame of reference as a tool for improving performance prediction, narrowing item specificity to one social role has also shown beneficial effects for predicting other context-related and general variables. Bowling and Burns (2010) compared self-reported data from workers employed in a variety of different organizations when answering general or at-work altered items from the IPIP Big Five Factor Markers. Work-related personality was a better predictor of job-related variables – job satisfaction and turnover intention were related with extraversion, agreeableness and neuroticism, absenteeism with agreeableness and work frustration with neuroticism. In several cases work-specific personality predicted incremental variance in these criteria after general personality was controlled for (similar to the findings of Bing et al., 2004, who studied the incremental validity of academic personality over the general one in predicting GPA). Furthermore, work-specific personality was a better predictor of several general criteria, such as physical health (which is related to extraversion), psychological health, and life satisfaction (both related to extraversion and agreeableness). Mlinarič (2012) found MINI IPIP Conscientiousness score and all six facets of the IPIP-NEO 300 Conscientiousness dimension, except the Deliberation facet, to be better predictors of the frequency of class attendance when the items were modified to the school-specific context compared to general personality measures. Heller and his co-workers

(2009) found that work personality, i.e., when items were answered according to work situations, was a better predictor of job satisfaction than global and home personality, whereas home personality was a better predictor of marital satisfaction compared to general and work personality. Heller and his co-workers (2009) and Bowling and Burns (2010) also found evidence that work-personality dimensions mediate the association between their corresponding global-personality dimensions and the relevant criterion variables.

The listed studies examined different aspects of the specified context on personality measurement. In general it is evident that psychometric properties of personality measures differ according to item specificity in the form of given specific social roles or contexts (e.g., school, work, home, general), although there are strong positive correlations between specific- and general-personality measures (e.g., Bowling & Burns, 2010; Heller et al., 2009; Reddock et al., 2011). These correlations support the assumption that personality is generally relatively stable across situations, but also show that different personalities do not completely overlap. In most cases, the original factor structures can describe situation-based personality measures adequately enough (e.g., Schmit et al., 1995).

*Topics to be considered when examining the frame-of-reference effect*

There are various limitations to the generalizability of the above-mentioned findings. One of the limitations is the use of student samples in some studies (their motivation to obtain a course credit may interfere with their responses). Some studies involved voluntary respondents in relatively low stakes situations (e.g., simulated selection contexts), that are not characteristic for a real selection process. Furthermore, simultaneous (or close-in-time) administration of the general and contextualized version of the scale may have led to a carry-over effect.

The reported findings were also not consistent over different social roles. Some of the more evident are the inconsistencies related to the Big Five dimensions compared to the facet level of the mostly measured conscientiousness. There is a reason to suspect that the frame-of-reference effect could be different for each facet and consequently result in different overall dimension scores. Some dimensions and facets may be more generalizable across situations. The inconsistent inclusion of the openness dimension in frame-of-reference studies leaves some open questions about how the dimension varies across situations. It is also not sure whether the same contextualization effects would be found for other personality constructs. Different effects of contextualization could be found with different personality measures. The effects of contextualization could also depend on

some characteristics of the instruments used (such as length, form, etc.).

In future, the effects of specific frame-of-reference instructions (as compared to the generic instructions) should be studied in more detail, as it is still not entirely clear what effect such a contextualization has on the predictive validity of the instrument. Although a decrease in the within-person variability has been considered as one of the potential effects of using specific frame-of-reference instructions (Lievens et al., 2008), some studies found mixed support for this thesis (e.g., Bowling & Burns, 2010). Furthermore, the effects of contextualizing each item separately vs. adding specific frame-of-reference instructions at the beginning of the instruments should be compared.

Schmit and his co-workers (1995) noted that it may be possible that items from different scales require different levels of context specificity to increase validity. It would therefore be of potential benefit for economizing the instruments to detect how much contextualization is needed – should we contextualize some items, most items, or every item. It would also be beneficial to know how specific the contextualization should be to achieve a desired increase in validity. When examining the effect size of the contextualization, one should also have in mind that the validities for predicting different criteria may not be directly comparable. For example, in predicting job performance with context-related personality, one may examine different criteria – e.g., colleague and employee ratings – and the effect of item contextualization on the validity for predicting both measures may be different. Therefore it is difficult to make generalizations about the contextualization effects over studies using different criteria. Differential increments in the predictive validity for different criteria may, nevertheless, be useful for understanding the relationships between various constructs.

When using contextualized instruments in practice, one must be aware that context-related measures may not increase the validity when the context is relatively unknown to the participant. Questionnaires imposing a specific role should therefore be carefully chosen when they are used, for example, in the selection process.

Last but not least, it should be mentioned that there may be individual differences in susceptibility to the contextualization effects. Schmit and his co-workers (1995) stated that an additional measure to be included in the studies on the contextualization effects would be a self-monitoring estimate. Namely, self-monitors are highly skilled at modifying their behaviour to social demands of a situation. Therefore, self-monitors could be more susceptible to contextual cues in personality measurement. Future studies should examine individual differences like these and also the key aspects of social situations that affect individuals to adopt a particular social role and give responses that are situation-specific and deviate from the ones that would describe their general personality.

## REFERENCES

- Augustine, A., & Larsen, J. R. (2012). Is a trait really the mean of states? Similarities and differences between traditional and aggregate assessments of personality. *Journal of Individual Differences*, 33(3), 131–137.
- Bing, M., Whanher, J. C., Davidson, H. K., & VanHook, J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology*, 89(1), 150–157.
- Bowling, N., & Burns, G. (2010). A comparison of work-specific and general personality measures as predictors of work and non-work criteria. *Personality and Individual Differences*, 49, 95–101.
- Chamorro-Premuzic, T., & Furnham, A. (2003). Personality traits and academic examination performance. *European Journal of Personality*, 17, 237–250.
- De Fruyt, F., & Mervielde, I. (1996). Personality and interests as predictors of streaming and achievement. *European Journal of Personality*, 10, 405–425.
- Gordon, E. D., & Holden, R. R. (1996). Use of item ratings to examine personality test item cognitive response processes. *Personality and Individual Differences*, 21, 897–905.
- Gordon, E. D., & Holden, R. R. (1998). Personality test item validity: Insights from “self” and “other” research and theory. *Personality and Individual Differences*, 25, 103–117.
- Heller, D., Ferris, D. L., Brown, D., & Watson, D. (2009). The influence of work personality on job satisfaction: Incremental validity and mediation effects. *Journal of Personality*, 77, 1051–1084.
- Hunthausen, J. M., Truxillo, D. M., Bauer, T. B., & Hammer, B. L. (2003). A field study of the frame-of-reference effect on personality test validity. *Journal of Applied Psychology*, 88, 545–551.
- Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, 93(2), 268–279.
- Mlinarič, V. (2012). *Učinek kontekstualiziranja postavk v osebnostnih vprašalnikih* [The effect of contextualizing items in personality questionnaires] (Unpublished undergraduate thesis). Department of Psychology, University of Ljubljana, Slovenia.
- Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, 66, 116–130.
- Reddock, C. M., Biderman, M. D., & Nguyen, N. T. (2011). The relationship of reliability and validity of personality tests to frame-of-reference instructions and within-person inconsistency. *International Journal of Selection and Assessment*, 19(2), 119–131.
- Robie, C., Schmit, M. J., Ryan, A. M., & Zickar, M. J. (2000). Effects of item context specificity on the measurement equivalence of a personality inventory. *Organizational Research Methods*, 3, 348–365.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, S. L. (1995). Frame-of-reference effects on personality scores and criterion-related validity. *Journal of Applied Psychology*, 80, 607–620.
- Wright, J. C., & Mischel, W. (1987). A conditional approach to dispositional constructs: The local predictability of social behavior. *Journal of Personality and Social Psychology*, 53(6), 1159–1177.