
01 Jan 2024

How to Frame the Frame of Reference: A Comparison of Contextualization Methods

Ann E. Schlotzhauer

Matthew A. Ng

Missouri University of Science and Technology, matthew.ng@mst.edu

Shiyang Su

Follow this and additional works at: https://scholarsmine.mst.edu/psysci_facwork



Part of the [Psychology Commons](#)

Recommended Citation

Schlotzhauer, A. E., Ng, M. A., & Su, S. (2024). How to Frame the Frame of Reference: A Comparison of Contextualization Methods. *Journal of Business and Psychology* Springer.

The definitive version is available at <https://doi.org/10.1007/s10869-024-09953-8>

This Article - Journal is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Psychological Science Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.



How to Frame the Frame of Reference: A Comparison of Contextualization Methods

Ann E. Schlotzhauer¹ · Matthew A. Ng¹ · Shiyang Su¹

Accepted: 29 April 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Personality measures are popular and useful in employment selection and academic contexts; however, concerns have been voiced regarding the strength of their association with desirable criteria. Contextualization (i.e., modifying measures to reflect the desired frame of reference, like work or school) has emerged as a promising option. Research has demonstrated that contextualizing personality measures increases predictive validity and enhances participants' perceptions of the assessments. However, few studies have compared contextualization methods to one another and, to date, only one study has compared the two most common forms of contextualization (i.e., instruction and tag contextualization), returning inconsistent findings. In a within-person, multi-wave study using a working sample ($N = 399$), we compared the relative efficacy of personality measures that are contextualized through manipulating the instructions and those contextualized through the addition of contextual item tags. We specifically contextualized the big five personality factors in order to predict work-related outcomes (i.e., job satisfaction, perpetrated incivility, job performance, creative job performance, and emotional exhaustion). Our study supports the use of tag-level contextualization and provides guidance on how to best implement contextual tags. Best practices, implications, and future research directions are discussed.

Keywords Personality · Contextualization · Frame-of-reference effect · Big five · Criterion-related validity

Although research has demonstrated the predictive validity of personality, and especially conscientiousness, in employment selection contexts, personality inventories remain a weaker predictor of job performance than cognitive ability tests (Barrick & Mount, 1991; Furnham et al., 2009; Guion & Gottier, 1965; Sackett et al., 2021; Tett et al., 1991). As such, there has been interest in identifying methods of increasing the predictive validity of personality measures, with contextualization emerging as a promising option (Schmit et al., 1995). Contextualization offers an easy and low-cost means of improving the prediction of desired outcomes (e.g., job performance) by limiting a test-taker's frame of reference to the desired context (e.g., their conscientiousness in work settings).

To date, research mostly provides evidence of the utility of contextualization as a whole, by comparing contextualized measures to non-contextualized measures. However, very few investigations have directly compared contextualization methods to one another and only one study has directly compared the efficacy of the two simplest and most popular forms of contextualization, instruction-level contextualization (i.e., modifying instructions to evoke the desired frame of reference) and tag-level contextualization (i.e., adding contextual tags to items). Although both of these contextualization methods have benefits compared to non-contextualized personality measures and are less resource-intensive than complete contextualization (i.e., fully rewriting items in the intended frame of reference mirroring common scale development practices including validation efforts with subject matter experts), researchers and practitioners currently have little guidance regarding which of these two methods should be preferred. We address this gap by comparing the utility of instruction-level and tag-level contextualized personality measures for predicting context-specific outcomes.

The present research boasts several strengths. Because most contextualization research has not directly compared

Additional supplementary materials may be found here by searching on article title <https://osf.io/collections/jbp/discover>.

✉ Ann E. Schlotzhauer
ann.schlotzhauer@ucf.edu

¹ Department of Psychology, University of Central Florida, Orlando, FL 32816, USA

instruction- and tag-level contextualization, this study represents an important contribution. Further, as illustrated in Table 1, relevant past research is marked by a variety of different design factors and inconsistent findings, which limit the conclusions that could be drawn regarding instruction- and tag-level contextualization. For example, several studies have contextualized personality to the school setting, using either instructions or tags, and presented correlations between conscientiousness and grade point average (GPA) (e.g., Bing et al., 2004; Lievens et al., 2008; Reddock et al., 2011; Schmit et al., 1995). However, a closer look at these studies does not provide a consistent picture supporting either instruction- or tag-level contextualization over the other method. Although one study has directly compared instruction- and tag-level contextualization (i.e., Swift & Peterson, 2019), generalizations from their findings (i.e., tags outperformed instruction-contextualization in one of

four comparisons) may be impeded by their cross-sectional design and limited predictor-criterion relationships.

Therefore, our comparison of instruction-level and tag-level contextualization advances the literature and provides useful guidance to researchers and practitioners alike. In addition, we utilize a diverse working sample and leverage a time-lagged, within-person design that is more methodologically rigorous than most contextualization research. By utilizing a time-lagged design rather than a cross-sectional design, we are better able to control for potential carryover effects associated with taking multiple personality assessment forms in one sitting as well as reducing the influence of common method variance (Podsakoff et al., 2003). In doing so, we increase confidence in the replicability of previous results regarding the general efficacy of contextualization. By utilizing a within-person design as opposed to a between-person design, we are able to utilize the most

Table 1 Brief overview of methodological contextualization studies

Citation	Methods compared	Context used	Analyses employed	Study design	Conclusions
Bing et al. (2004)	Base and tag	School	Hierarchical regression	Within subjects, cross-sectional	Tag > base
Bing et al. (2014)	Base and complete	Work	Correlation comparison and hierarchical regression	Within subjects, cross-sectional	Complete > base
Bowling and Burns (2010)	Base and tag	Work	Correlation comparison and hierarchical regression	Within subjects, cross-sectional	Tag > base
Heller et al. (2009)	Study 1: base and instruction	Work and home	Hierarchical moderated regression	Between subjects, cross-sectional	Instruction > base
Holtrop et al. (2014)	Base, tag, and complete	School	Hierarchical regression	Within subjects, 2 time points	Complete > tag > base
Hunthausen et al. (2003)	Base and instruction	Work	Correlation comparison and moderated regression	Between subjects, cross-sectional	Instruction > base
Lievens et al. (2008)	Study 1: base and instruction	School	Correlation comparison	Between subjects, cross-sectional	Instruction > base
Pace and Brannick (2010)	Base and complete	Work	Hierarchical regression	Between subjects, cross-sectional	Complete > base
Reddock et al. (2011)	Base and tag	School	Correlation comparison	Within subjects, cross-sectional	Tag > base
Robie and Risavy (2016)	Base and tag	School	Hierarchical regression	Between subjects, cross-sectional	No significant improvement
Robie et al. (2017)	Base, tag, and complete	School	Hierarchical regression	Within subjects, 3 time points	Tag and complete > base
Schmit et al. (1995)	Studies 1 and 2: base and tag	Work and school	ANOVA and correlation comparison	Within subjects, cross-sectional	Tag > base
Swift and Peterson (2019)	Studies 1 and 2: base and tag; study 3: base, instruction, and tag	Work, school, and romantic	K-fold regression	Within subjects, cross-sectional	Tag > base; instruction had no significant improvement over base

Base refers to non-contextualized personality scales. The Conclusions column summarizes the majority of findings reported in a given study. Holtrop et al. (2014) employed a time lag in their design with the base version of the personality scales and one of either the tag or complete contextualized versions delivered at time 1 and the other contextualized version and outcomes collected at time 2. Only study 1 of Heller et al. (2009) and Lievens et al. (2008) are reported, as their respective second studies did not include relevant comparisons

common analyses in similar methodological comparisons in the literature (i.e., correlation comparison and hierarchical regression); see Table 1. In addition to these common analyses, we employ relative weights analysis (RWA) to account for the high multi-collinearity expected in a hierarchical linear regression comparing the separate methods (Tonidandel & LeBreton, 2015). We examine work-relevant outcomes associated with each of the big five personality factors, rather than focusing solely on conscientiousness as a predictor. Nonetheless, a comparison of instruction- and tag-level contextualization may leave readers wondering if they should simply leverage both. In some situations, this may be appropriate. However, there may be scenarios in which researchers and practitioners are wary of altering a scale's instructions and potentially distracting from other meaningful aspects of the instructions (e.g., a time frame, a target individual).

In sum, the current manuscript aims to apply rigorous methodology to (a) replicate past findings regarding the efficacy of contextualization generally and (b) compare the efficacy of the two most common and accessible contextualization methods.

Background

Contextualization of Personality

For many decades, a growing contingent of psychologists have challenged the traditional trait approach to understanding human personality (Mischel, 1973). Specifically, the cognitive-affective system theory of personality suggests that personality is conditional on situational cues as well as individual factors such as a person's life experience; as such, behavior can only be expected to be consistent across situations to the extent that those situations elicit similar feelings, motives, and interpretations (Mischel & Shoda, 1995). For example, a person might be more consistently agreeable and extraverted amongst friends than they are at work. Drawing from this thinking, contextualized personality measures strive to limit a test-taker to a specific context as they answer items about their characteristic ways of thinking and behaving (Bing et al., 2004). For example, if a personality measure is being used to predict job performance, the test-taker would be encouraged to answer items according to how they act in work settings. Thus, contextualized measures provide a frame of reference in order to assess personality within a given context, rather than general personality.

In support of this thinking, Lievens and colleagues (2008) explored how contextualization increases criterion-related validity, finding that contextualization reduces both between-person variability and within-person inconsistency. In other words, contextualization reduces the likelihood of

two different test-takers considering different life domains (e.g., work and social settings) in responding to the same personality inventory. It also reduces the likelihood of each test-taker drawing on different life domains in responding to individual items on the inventory. Beyond these explanations, the symmetry principle has also been leveraged to explain how contextualization improves criterion-related validity. In short, the symmetry principle suggests that when the specificity of a predictor and outcome are matched or symmetrical, their relationship is stronger than when the specificity is unmatched or asymmetrical (Ajzen, 2005; Schulze et al., 2021).

The formal study of contextualized personality measures dates back only a few decades (Schmit et al., 1995). However, in this time, a number of studies have examined the relationship between contextualized personality and work outcomes (Bing et al., 2014; Bowling & Burns, 2010; Hunthausen et al., 2003; Pace & Brannick, 2010; Robie et al., 2000, 2001; Schmit et al., 1995; Swift & Peterson, 2019). Meta-analytic findings suggest that validity for predicting supervisor-rated job performance can be improved for all of the big five personality factors through contextualization (Shaffer & Postlethwaite, 2012). Overall, Shaffer and Postlethwaite (2012) found a mean validity increase from 0.11 for non-contextualized personality measures to 0.24 for contextualized personality measures.

Importantly, contextualized personality measures also boast other advantages over non-contextualized personality measures. In selection contexts, non-contextualized personality measures are often viewed as less job-relevant than other assessments (Hausknecht et al., 2004). As a result, applicants respond less favorably to non-contextualized personality measures than they do to interviews, work samples, or even cognitive ability tests (Hausknecht et al., 2004). Contextualizing personality inventories increases their face validity and perceived predictive validity (Holtrop et al., 2014; Robie et al., 2017), improving the test-taker experience (Ployhart et al., 2003). This is a worthy goal as face validity and perceived predictive validity are associated with a variety of applicant reactions to the selection process, including perceptions of justice and intentions to accept job offers (Hausknecht et al., 2004). Thus, the contextualization of personality measures can not only increase predictive validity, allowing organizations to make better selection decisions, but also improve the applicant experience, making applicants more likely to accept employment offers.

Methods of Contextualization

Although a great deal has been learned about the value of contextualized personality measures over the last several decades, there is much more left to uncover related to how contextualization should be approached. In past research,

personality items have been contextualized through one of three methods (Holtrop et al., 2014). The first method, instruction-level contextualization, manipulates a scale's instructions to direct individuals to answer items according to how they typically behave in a given context (e.g., "To what extent do the following items reflect your tendencies at work?"). The second method, tag-level contextualization, adds contextual tags (e.g., "at work," "at school") to the individual items of the scale. The third method, complete contextualization, provides a frame of reference by fully rewriting the items of the scale to align with the target context (Holtrop et al., 2014). For example, a generic emotional stability item reading "I remain calm during emergencies" could be replaced by a work-contextualized item reading "I handle pressing work tasks with steady nerves" (Bing et al., 2014, p. 171).

Importantly, researchers have utilized all three of these methods to effectively match the frame of reference of personality measures to that of relevant outcomes. Evidence suggests that instruction-level contextualized personality measures better predict relevant outcomes than non-contextualized personality measures (Hunthausen et al., 2003; Lievens et al., 2008). Tag-level contextualized personality is also more effective than non-contextualized personality (Bing et al., 2004; Bowling & Burns, 2010; Reddock et al., 2011). Finally, complete contextualization also seems to increase predictive validity of personality measures compared to non-contextualized ones (Bing et al., 2014; Pace & Brannick, 2010; Swift & Peterson, 2019). Overall, meta-analytic findings support the conclusion that contextualized measures are more valid predictors of context-relevant outcomes than are non-contextualized measures (Shaffer & Postlethwaite, 2012).

While all three contextualization methods can improve the predictive validity of personality, the extant literature provides little guidance on which contextualization method to prefer. In fact, only a few studies have compared contextualization methods to one another. A study by Holtrop and colleagues (2014) demonstrated that personality measures with both tag-level and complete contextualizations tended to explain more variance in criteria than did non-contextualized personality measures; further, completely contextualized scales tended to outperform tag-contextualized scales. Robie and colleagues (2017) reported that tag-contextualized and completely contextualized personality measures generally outperformed non-contextualized measures, but only found partial support for the advantage of complete contextualization over tag-level contextualization. The one study to date that has compared instruction- and tag-level contextualization reported inconclusive findings from their cross-sectional study, with tag contextualization outperforming instruction contextualization in one comparison and no significant difference in the other three comparisons (Swift

& Peterson, 2019). Of note, the instruction-level contextualization employed in the study only outperformed base personality in one of the four comparisons (Swift & Peterson, 2019).

Although there is evidence that completely rewriting items to fit the target context can produce strong results (Holtrop et al., 2014; Robie et al., 2017), there are also serious disadvantages to complete contextualization. First, the process is far more time- and labor-intensive than instruction- or tag-level contextualization. Complete contextualization requires a lengthy process, such as "(1) generating examples, (2) developing a preliminary list of items, (3) back-translation, (4) revision, and (5) a final check by two experts on personality ratings who assigned the completely contextualized items to the facet scales used in the inventory" (Robie et al., 2017, p. 59). Pace and Brannick (2010) described similar efforts, including an additional data collection to aid in development and validation of the contextualized scale. Even with these efforts, additional concerns have been expressed about the effects of contextualization on the psychometric properties of scales (Robie & Risavy, 2016). Specifically, complete contextualization may alter the content so extensively that the items no longer represent the intended domain; validation efforts are rarely pursued and, when they are, questions remain about whether they sufficiently address such concerns (see Heggstad et al., 2019). In short, complete contextualization constitutes the creation of a new scale, requiring extensive validation efforts to ensure psychometric validity. As a result, instruction-level and tag-level contextualizations have been significantly more common in the research literature and represent more accessible options for practitioners (Holtrop et al., 2014). For an overview of previous methodological approaches to contextualization research, refer to Table 1.

The purpose of the current study is twofold. We first aim to replicate previous findings regarding the general efficacy of contextualized personality measures with a more rigorous study design than has typically been utilized in contextualization research. Previous studies mostly used cross-sectional designs. We leverage a within-person, time-lagged design using a diverse, working sample. Using the multi-wave design and assessing lagged relationships with temporally separated variables should limit the risk of common method variance, which may have altered the nature of relationships observed in past research (Johnson et al., 2011; Podsakoff et al., 2003, 2012). Our design also limits the risk of potential carryover effects or demand characteristics that may occur when participants respond to contextualized and non-contextualized personality measures back-to-back (e.g., Schmit et al., 1995). With this more rigorous study design, we expect that, regardless of contextualization method, contextualized measures will outperform non-contextualized measures. We assess

the efficacy of contextualization utilizing three analyses in order to explore (a) the relationships between personality and various work outcomes by comparing correlational strengths, (b) the predictive ability of the measures via hierarchical regression, and (c) the contribution of each method in explaining variance in the outcome with relative weights analysis.

Hypothesis 1: Personality scales employing any form of contextualization (i.e., instruction or tag) will outperform non-contextualized personality scales in terms of the (1a) strengths of correlations with criteria, (1b) incremental predictive validity, and (1c) relative importance of the predictors.

Further, we seek to extend current knowledge in this field by directly comparing the two most common and accessible methods of contextualization: instruction-level and tag-level contextualization. There is limited empirical foundation for hypothesizing that one form of contextualization will outperform the other, with the only direct comparison finding no significant difference between instruction- and tag-level contextualization in the majority of their comparisons (Swift & Peterson, 2019). However, there is some theoretical rationale to prefer tag-level contextualization. As mentioned, the theoretical foundation for contextualizing personality measures has typically been the cognitive-affective system theory of personality, which suggests that personality is conditional on situational cues (Mischel & Shoda, 1995). As a result, this theory posits that behavior will be consistent across situations to the extent that those situations present similar cues (Mischel & Shoda, 1995). Because tag-level contextualization repeats the frame of reference in each item, it is reasonable to expect that it will more successfully reinforce the proper context than instruction-level contextualization, which lists the target context only once. Further, previous research on survey methodology suggests that participants are less likely to read instructions than item text (Oppenheimer et al., 2009; Shamon & Berning, 2020). As a result, researchers have suggested that critical information should be provided in items, rather than instructions, whenever possible (Shamon & Berning, 2020). Thus, we expect that the repetition of the desired frame of reference in each item should lead tag-level contextualization to outperform instruction-level contextualization.

Hypothesis 2: Tag-level contextualized personality scales will outperform instruction-level contextualized personality scales in terms of the (2a) strengths of correlations with criteria, (2b) incremental predictive validity, and (2c) relative importance of the predictors.

Pilot Study Method

First, a pilot study was conducted to explore the placement of contextual tags within items. Just as past research provides little guidance for whether instruction- or tag-contextualization should be preferred, there is also no research guiding how tag-level contextualization should be implemented. Past research utilizing tag-level contextualization has typically added tags to the ends of items (e.g., Bowling & Burns, 2010; Holtrop et al., 2014; Lievens et al., 2008; Robie et al., 2017) or a mix of beginnings and ends of items (e.g., Holtz et al., 2005; Reddock et al., 2011; Robie et al., 2000; Schmit et al., 1995). To our knowledge, however, no researchers have provided rationale for the locations of these contextual tags beyond basing their decisions on grammatical fit (e.g., Holtrop et al., 2014; Robie et al., 2017).

The location of item tags deserves greater attention. When individuals are presented with information to memorize, both primacy (i.e., items early in the list are remembered better) and recency effects (i.e., items late in the list are remembered better) are reliably demonstrated (Kelley et al., 2013). Considering the cognitive effects associated with primacy and recency, tag location may impact the salience of the frame of reference in working memory. Greater salience could lead individuals to more carefully consider their behavior in the target context, leading to greater criterion-related validity. Importantly, evidence suggests that response-order effects can have meaningful impacts on survey research (Holbrook et al., 2007; Krosnick & Alwin, 1987). Due to limited theoretical rationale or empirical evidence about tag locations, we explored the effects of three possible locations (i.e., beginnings of items, ends of items, or split between beginnings and ends of items) in the pilot study in order to inform tag usage in our primary study.

Sample and Procedure

The pilot study contextualized conscientiousness to an academic setting. Specifically, data were collected in the spring of 2020 from undergraduate students at a large university in the USA. This study focused on the relationship between conscientiousness and academic performance, operationalized as college grade point average (GPA). Data were removed for any participant who was unable to provide an official college GPA. Overall, the sample ($N = 257$) was 55.6% female and had an average GPA of 3.44 ($SD = 0.46$).

Participants were randomly assigned to one of four conditions. In the base condition, participants completed

the conscientiousness scale described below without any manipulations; the instructions prompted them, “To what extent do the following items reflect your tendencies?” For each of the three tag conditions, the participants viewed the same instructions as the base condition, but each item of the scale had the words “at school” added to the beginning or the end of the item. The three tag conditions will be referred to as *primacy* (in which the “at school” tag is added to the beginning of each item), *recency* (in which the “at school” tag is added to the end of each item), and *split* (in which the “at school” tag alternates between the beginning and end of each item). See Appendix A for the contextualized items. Descriptive information across each of the conditions can be found in Table 2.

Measures

Conscientiousness

The 20-item Big-Five Factor Markers conscientiousness scale from the International Personality Item Pool was used as the general measure of conscientiousness (Goldberg, 1992). Sample items include “I am always prepared” and “I like order.” Responses were recorded on a seven-point Likert scale (1 = Strongly Disagree; 7 = Strongly Agree). See Table 3 for internal consistency across conditions.

Academic Performance

The criterion variable of academic performance was operationalized as participants’ grade point averages (GPAs). Participants were asked to self-report their current college GPA at the end of the survey. Evidence suggests that undergraduate research participants provide highly accurate self-reports of their college GPAs (Caskie et al., 2014; Cassady, 2000; Gray & Watson, 2002; Nofle & Robins, 2007).

Table 3 Correlations and scale statistics in pilot study

Condition	Correlation with GPA	<i>p</i> -value	Coefficient α
Base	-.05	.715	.91
Primacy Tag	.19	.136	.93
Recency Tag	.10	.438	.90
Split Tag	.27	.030	.92
Overall	.14	.025	

Pilot Study Results

Correlations were calculated to examine relationships between the different versions of the conscientiousness measure and academic performance; see Table 3. Collapsed across conditions, conscientiousness correlated positively with GPA. This correlation aligns with meta-analytic findings on the relationship between the Big-Five Factor Markers conscientiousness scale and GPA (McAbee & Oswald, 2013). Interestingly, the correlation between conscientiousness and GPA was significant in the split tag condition, but not in the primacy or recency conditions. Following the *z*-test method for comparing correlations from independent samples (Cohen et al., 2013; Eid et al., 2011), we calculated the difference between the strengths of the correlations. The correlation between conscientiousness and GPA in the split condition was neither significantly stronger than that of the primacy condition ($z = 0.49, p = 0.311$), nor that of the recency condition ($z = 1.00, p = 0.158$). The split condition did, however, evidence a significantly stronger correlation between conscientiousness and GPA than that of the non-contextualized condition ($z = 1.80, p = 0.036$), which was not the case for the other tag conditions. Note that the above *p*-values represent one-tailed tests.

Table 2 Descriptives across the pilot study conditions

Condition	<i>N</i>	Gender			Mean GPA (<i>SD</i>)	Mean conscientiousness (<i>SD</i>)	Cohen’s <i>d</i> ^a
		Man	Woman	Other			
Base	65	22	42	1	3.39 (0.47)	4.82 (0.90)	
Primacy tag	64	30	32	2	3.50 (0.46)	5.38 (0.92)	0.62
Recency tag	65	29	35	1	3.47 (0.43)	5.49 (0.80)	0.79
Split tag	63	29	34	0	3.41 (0.46)	5.49 (0.90)	0.74
Overall	257	110	143	4	3.44 (0.46)	5.29 (0.92)	

^aThis column reflects the Cohen’s *d* effect size of the difference between the mean conscientiousness in each condition compared to the mean in the Base condition

Pilot Study Discussion

The preliminary findings suggest that split tag contextualization should be preferred over primacy or recency tags. There are at least two potential explanations for why a split approach toward tag contextualization may be more effective. First, the flexibility of utilizing tags either in the beginning or end allows for researchers to better conform to grammatical conventions, thus reducing the potential cognitive load for participants. Additionally, providing contextual information at different locations within each item introduces more variety, potentially staving off cognitive shortcuts participants may engage in when reading items. In other words, when the same phrase is consistently repeated in the same location of each item, participants may be more likely to ignore this phrase and only attend to the unique parts of each item. Based on the pilot study, the split tags were then used as the form of tag contextualization in the primary study.

An important note regarding the pilot study is that these data were collected between January 28 and April 10, 2020. In response to the COVID-19 pandemic, all courses at the institution attended by this student sample were moved online beginning Monday, March 16, 2020. We expect that this salient disruption may be partially responsible for the rather low correlation observed between conscientiousness and GPA in the base condition. For those participants responding to a non-contextualized conscientiousness inventory during the early days of COVID-19, the pandemic may have been the most salient context available and participants may have self-imposed this context in their responses. Although the unique timing of our data collection likely limits the generalizability of the correlation between base conscientiousness and GPA, it also emphasizes the importance of contextualization, especially in situations when a different, undesired context may be more salient than the desired context. Further highlighting the importance of contextualization, the mean conscientiousness observed in the base condition was notably lower than the mean conscientiousness observed in each of the contextualized conditions (see Table 2 for effect sizes).

Primary Study Method

The primary study compared instruction-contextualized, tag-contextualized, and non-contextualized personality within a working sample. A multi-wave data collection allowed us to collect three versions of the personality measures with minimal potential spillover effects between conditions and provide a lagged test of their predictive

validity. This study examined the predictive validity of each of the big five personality traits in predicting established criteria associated with the work context. Specifically, extraversion has been shown to relate to job satisfaction (Judge et al., 2002), agreeableness to perpetrated incivility and aggression (Taylor & Kluemper, 2012; Welbourne et al., 2020), conscientiousness to job performance (Dudley et al., 2006), openness to experience to creative job performance (Pace & Brannick, 2010; Schilpzand et al., 2011), and neuroticism to emotional exhaustion (Kammeyer-Mueller et al., 2016; Sosnowska et al., 2019).

Data were collected in four waves on Amazon Mechanical Turk (MTurk) using the CloudResearch Toolkit in the spring of 2021. To ensure high-quality data, our study was only available to CloudResearch-approved participants who had completed more than 10,000 MTurk human intelligence tasks (HITs) with an overall approval rating higher than 98%. We also followed MTurk best practice recommendations for increasing data quality, including the use of captcha verification and attention checks (Aguinis et al., 2021). Participants were screened for eligibility (i.e., working on average 35 h per week or more, employed at their current organization for at least three months, not self-employed, at least 21 years of age, residing in the USA, interacting with coworkers and/or supervisors on a weekly basis) and then compensated for the successful completion of each of four surveys. Each survey was administered 1 month apart with personality measures collected at the first, second, and third time points and the outcome measures collected at the fourth time point. To reduce participant fatigue and potential spillover effects between conditions and to encourage more thoughtful responses, participants responded to only one version of the personality measure at each of the first three time points. To control for potential order effects, the order of administration (e.g., base then instruction then tag; tag then base then instruction) was randomized and counterbalanced across participants. In total, there were six orders of administration; each was completed by 60–74 participants of the final retained sample. All outcomes (i.e., job satisfaction, perpetrated incivility, job performance, creative job performance, and emotional exhaustion) were measured at the fourth and final time point.

Participants

A total of 534 participants met the eligibility requirements and completed the wave 1 survey. In total, 465 participants completed the wave 2 survey, 438 completed the wave 3 survey, and 406 completed the wave 4 survey. Thus, 76.0% of those who completed the first survey also completed the final survey. Two participants were removed from the analytic sample for failing two attention checks at wave 1, and one participant was removed for failing two attention checks

at wave 2. Participants who did not have complete data for all variables of interest or whose data could not be matched across waves (i.e., due to a missing ID) were excluded from analyses. The final analytic sample consisted of 399 participants who ranged in age from 23 to 78 years of age ($M = 43.2$, $SD = 11.7$). Roughly half (53.4%) of the sample identified as female. The majority (82.2%) identified as Caucasian or white, with the next largest group (7.0%) identifying as Asian or Asian American. Participants were employed in a variety of industries with education (13.3%), health care or social assistance (12.3%), and professional, scientific, or technical services (11.5%) most highly represented.

Independent samples t -tests and crosstabs with pairwise z -tests using Bonferroni-corrected p -values were used to compare participants included in the analytic sample to those who completed the wave 1 survey but were not included in the final sample. Participants included in the final sample did not differ significantly from those excluded after the wave 1 survey in terms of age, race, or gender identity. Compared to those who were not included in the final sample, participants in the analytic sample were less likely to report working in the Information industry (8.3% vs. 14.1%). There were no other differences in industry composition. We also compared the personality measures completed on the wave 1 survey by these two groups. The average score on the tag-contextualized conscientiousness scale was higher for those included in the final sample ($M = 4.31$, $SD = 0.60$) compared to those who were not included ($M = 4.06$, $SD = 0.71$, $t = -2.22$, $p = 0.03$). The other 14 personality comparisons yielded non-significant differences between groups.

Measures

Base Personality

We assessed non-contextualized extraversion, agreeableness, conscientiousness, openness to experience, and neuroticism using the mini-IPIP (Donnellan et al., 2006). The scale consists of four items assessing each personality trait, for a total of 20 items. Minor edits were made to two items to remove a specific frame of reference. Specifically, an extraversion item reading “I talk to a lot of different people at parties” was edited to read “I talk to a lot of different people.” Similarly, a conscientiousness item reading “I get chores done right away” was edited to read “I get tasks done right away.” The instructions for this condition read, “To what extent do the following items reflect your tendencies?” Participants responded to all three personality measures on a scale from 1 (Strongly Disagree) to 5 (Strongly Agree). Internal consistency was acceptable for extraversion ($\alpha = 0.90$), agreeableness ($\alpha = 0.85$), conscientiousness ($\alpha = 0.76$), openness to experience ($\alpha = 0.83$), and neuroticism ($\alpha = 0.81$).

Instruction-Contextualized Personality

We adapted the base personality measure to measure work-specific personality by manipulating the instructions. Participants read the following instructions before responding to items: “To what extent do the following items reflect your tendencies **at work**?” Other than this change to instructions, the items were the same as those in the base condition. Internal consistency was acceptable for extraversion ($\alpha = 0.88$), agreeableness ($\alpha = 0.84$), conscientiousness ($\alpha = 0.79$), openness to experience ($\alpha = 0.82$), and neuroticism ($\alpha = 0.80$).

Tag-Contextualized Personality

Because split tags were the most effective form of tag contextualization in the pilot study, this study applied tag contextualization in the same manner. Specifically, the base personality measure was adapted by adding “at work” tags to either the beginning or end of an item. We applied the “at work” tags such that items made grammatical sense and items were presented in an alternating order (i.e., the first item started with the “at work” tag, the next item ended with the “at work” tag, and so on). See Appendix A for the items. The individual items were presented in the same order in all three conditions and items representing each trait were grouped together (e.g., four extraversion items followed by four agreeableness items, and so on). Similar to the base condition, the instructions for this condition read, “To what extent do the following items reflect your tendencies?” Internal consistency was acceptable for extraversion ($\alpha = 0.86$), agreeableness ($\alpha = 0.86$), conscientiousness ($\alpha = 0.73$), openness to experience ($\alpha = 0.81$), and neuroticism ($\alpha = 0.78$). See Appendix B for information about the factor structure of the three versions of the personality scale.

Job Satisfaction

Job satisfaction was assessed using the three-item subscale from the Michigan Organizational Assessment Questionnaire (Bowling & Hammond, 2008; Cammann et al., 1983) on a scale from 1 (Strongly Disagree) to 5 (Strongly Agree). A sample item includes “All in all I am satisfied with my job.” The measure evidenced high internal consistency ($\alpha = 0.91$).

Perpetrated Incivility

Participants were asked how frequently they had engaged in incivility toward their supervisor or coworkers over the past month (Cortina et al., 2001) on a scale from 1 (Never) to 5 (Always). A sample item includes “made demeaning or derogatory remarks about them.” The four-item scale demonstrated high internal consistency ($\alpha = 0.91$).

Job Performance

Participants were asked to recall their most recent performance evaluation and estimate how they were rated relative to their coworkers on five criteria (Pearce & Porter, 1986). A sample criterion reads “overall performance.” Participants indicated on a sliding scale what percentile (10th percentile–90th percentile, using increments of 10) they believed represented their relative performance on each criteria. The measure evidenced high internal consistency ($\alpha = 0.98$).

Creative Job Performance

Participants were asked to report the extent to which they produce original or novel work (Oldham & Cummings, 1996) on a scale from 1 (Strongly Disagree) to 5 (Strongly Agree). A sample item includes “The work I produce is creative.” These three items evidenced high internal consistency ($\alpha = 0.94$).

Emotional Exhaustion

Participants completed the emotional exhaustion subscale from the Maslach Burnout Inventory (Maslach & Jackson, 1981) on a scale from 1 (Strongly Disagree) to 5 (Strongly Agree). A sample item includes “I feel emotionally drained from my work.” This nine-item measure had high internal consistency ($\alpha = 0.96$).

Analytical Approach

In line with previous studies comparing contextualized and non-contextualized measures (Bing et al., 2004, 2014; Bowling & Burns, 2010; Pathki et al., 2022), we assess the strength of relationships through correlation and the incremental predictive validity of the measures through hierarchical regression. Further, we advance the literature by utilizing relative weights analysis to compare the relative importance of predictors.

Primary Study Results

Base, instruction-contextualized, and tag-contextualized personality measures were correlated with established outcomes over time. All correlations were in the expected directions, supporting the use of these established criteria. Specifically, extraversion, conscientiousness, openness to experience, and neuroticism correlated positively with job satisfaction, job performance, creative job performance,

and emotional exhaustion respectively; agreeableness correlated negatively with perpetrated incivility (see Table 4).

Hypothesis 1 posited that contextualized personality measures would outperform non-contextualized measures. Specifically, this hypothesis was assessed by (H1a) comparing correlational strengths through Steiger’s (1980) *z*-tests, (H1b) examining improvements in criterion prediction through hierarchical regression analyses, and (H1c) evaluating the relative contributions of predictors through relative weight analyses. First, correlations between base personality, contextualized personality, and their associated outcomes were compared using Steiger’s (1980) *z*-tests (see also Eid et al., 2011). See Table 5. The instruction-contextualized version of a personality trait measure demonstrated a significantly stronger association with its criterion than the base version in one case (i.e., openness to experience). The tag-contextualized version of a personality trait measure had a significantly stronger association with its criterion than the base version in three cases (i.e., extraversion, openness to experience, and neuroticism). No base personality measure demonstrated a significantly stronger association with its criterion than contextualized measures.

Hierarchical regression analyses were then used to evaluate the incremental predictive validity of contextualized measures over and above non-contextualized measures. Specifically, the base version of a personality trait measure (e.g., non-contextualized conscientiousness) was entered in step 1 predicting the associated outcome (e.g., job performance); then, the work-contextualized version of that personality trait measure (e.g., instruction- or tag-contextualized conscientiousness) was entered in step 2. A significant R^2 change (ΔR^2) indicates that the addition of the work-contextualized personality trait measure significantly improved prediction of the outcome over the base personality trait measure. Results can be found in Table 6. Instruction-level contextualization demonstrated incremental validity for associated outcomes for agreeableness, openness to experience, and neuroticism. Tag-level contextualization demonstrated incremental validity for associated outcomes for extraversion, agreeableness, openness to experience, and neuroticism.

To evaluate hypothesis 1c, relative weight analysis (RWA) was utilized to examine the relative importance or contribution of contextualized and non-contextualized personality trait measures toward the total predicted criterion variance. RWA is particularly appropriate for comparing the relative importance of predictors in situations where the predictors are correlated with one another (Tonidandel & LeBreton, 2015). Thus, five separate RWAs were conducted using RWA Web (Tonidandel & LeBreton, 2015). The three predictors (i.e., base, instruction-contextualized, and tag-contextualized versions of a personality trait) were included in one model for predicting each criterion. Table 7 reports the raw relative weights associated

Table 4 Descriptives, correlations, and scale reliabilities in primary study

Variable	Mean (SD)	1	2	3	4	5	6	7	8	9	10
1. B extraversion	2.71 (1.11)	(.90)									
2. I extraversion	2.77 (1.10)	.88***	(.88)								
3. T extraversion	2.91 (1.03)	.76***	.75***	(.86)							
4. B agreeableness	3.81 (0.90)	.27***	.22***	.34***	(.85)						
5. I agreeableness	3.81 (0.86)	.24***	.25***	.34***	.79***	(.84)					
6. T agreeableness	3.74 (0.91)	.25***	.27***	.43***	.77***	.79***	(.86)				
7. B conscientiousness	4.12 (0.74)	.15**	.17***	.16***	.26***	.26***	.27***	(.76)			
8. I conscientiousness	4.15 (0.75)	.10*	.16**	.13*	.19***	.26***	.25***	.78***	(.79)		
9. T conscientiousness	4.25 (0.67)	.07	.12*	.10*	.27***	.32***	.28***	.70***	.73***	(.73)	
10. B openness	3.82 (0.93)	.27***	.27***	.27***	.42***	.39***	.38***	.29***	.27***	.32***	(.83)
11. I openness	3.88 (0.88)	.27***	.29***	.28***	.40***	.43***	.40***	.28***	.32***	.30***	.84***
12. T openness	3.77 (0.88)	.34***	.37***	.39***	.44***	.44***	.47***	.34***	.36***	.38***	.78***
13. B neuroticism	2.27 (0.95)	-.30***	-.30***	-.31***	-.30***	-.26***	-.29***	-.46***	-.42***	-.35***	-.36***
14. I neuroticism	2.22 (0.90)	-.26***	-.30***	-.28***	-.29***	-.28***	-.29***	-.43***	-.43***	-.32***	-.36***
15. T neuroticism	2.10 (0.85)	-.27***	-.28***	-.34***	-.35***	-.34***	-.37***	-.40***	-.38***	-.41***	-.31***
16. Job satisfaction	3.80 (1.07)	.31***	.29***	.38***	.29***	.33***	.35***	.19***	.19***	.13*	.18***
17. Perpetrated incivility	1.37 (0.69)	.03	.04	-.06	-.32***	-.35***	-.33***	-.25***	-.22***	-.23***	-.22***
18. Job performance	7.45 (2.06)	.09	.10	.15**	.13**	.18***	.16**	.17***	.17***	.18***	.09
19. Creative performance	3.40 (1.10)	.32***	.33***	.37***	.29***	.34***	.35***	.23***	.21***	.18***	.30***
20. Emotional exhaustion	2.40 (1.12)	-.32***	-.30***	-.32***	-.24***	-.25***	-.24***	-.24***	-.23***	-.22***	-.23***
Variable	Mean (SD)	11	12	13	14	15	16	17	18	19	20
1. B extraversion	2.71 (1.11)										
2. I extraversion	2.77 (1.10)										
3. T extraversion	2.91 (1.03)										
4. B agreeableness	3.81 (0.90)										
5. I agreeableness	3.81 (0.86)										
6. T agreeableness	3.74 (0.91)										
7. B conscientiousness	4.12 (0.74)										
8. I conscientiousness	4.15 (0.75)										
9. T conscientiousness	4.25 (0.67)										
10. B openness	3.82 (0.93)										
11. I openness	3.88 (0.88)	(.82)									
12. T openness	3.77 (0.88)	.81***	(.81)								
13. B neuroticism	2.27 (0.95)	-.34***	-.40***	(.81)							
14. I neuroticism	2.22 (0.90)	-.41***	-.42***	.86***	(.80)						

Table 4 (continued)

Variable	Mean (SD)	11	12	13	14	15	16	17	18	19	20
15. T neuroticism	2.10 (0.85)	-.33***	-.43***	.72***	.75***	(.78)					
16. Job satisfaction	3.80 (1.07)	.22***	.28***	-.42***	-.42***	-.57***	(.91)				
17. Perpetrated incivility	1.37 (0.69)	-.19***	-.26***	.31***	.32***	.40***	-.30***	(.91)			
18. Job performance	7.45 (2.06)	.11*	.14**	-.16**	-.18***	-.23***	.22***	-.14**	(.98)		
19. Creative performance	3.40 (1.10)	.35***	.45***	-.29***	-.33***	-.36***	.39***	-.20***	.12*	(.94)	
20. Emotional exhaustion	2.40 (1.12)	-.22***	-.32***	.52***	.52***	.64***	-.67***	.37***	-.22***	-.30***	(.96)

B, I, and T indicate the base, instruction-contextualized, and tag-contextualized versions of a construct, respectively. Coefficient alphas are presented parenthetically along the diagonal. $N = 399$
 * $p < .05$; ** $p < .01$; *** $p < .001$

with each predictor (i.e., an additive decomposition of the model R^2) and their statistical significance based on bias corrected and accelerated confidence intervals (see Tonidandel et al., 2009), rescaled relative weights (i.e., the raw relative weights rescaled to reflect the percentage of predicted variance in the criterion that can be attributed to each predictor), and a comparison of the predictors (i.e., comparing the raw relative weights of the instruction-contextualized and tag-contextualized scales with the base scale). The tag-contextualized version of a scale predicted significantly more variance in the related criterion in three cases (i.e., extraversion, openness to experience, and neuroticism). The instruction-contextualized version of a scale did not predict significantly more variance in the related criterion in any cases. All in all, this series of analyses provided support for hypotheses 1a, 1b, and 1c.

Hypothesis 2 posited that tag-contextualized personality measures would outperform instruction-contextualized measures. Similar to hypothesis 1, this hypothesis was assessed using Steiger's (1980) z -tests (H2a), hierarchical regression analyses (H2b), and relative weight analyses (H2c). First, the correlations between instruction-contextualized and tag-contextualized personality measures with associated outcomes were compared using Steiger's (1980) z -tests (see also Eid et al., 2011). Results are described in Table 5. The tag-contextualized version of a personality trait measure had a significantly stronger association with its criterion than the instruction-contextualized version in three cases (i.e., extraversion, openness to experience, and neuroticism). These analyses are generally supportive of hypothesis 2a.

Next, hierarchical linear regression was leveraged to assess hypothesis 2b. The instruction-contextualized version of each personality measure (e.g., conscientiousness) was entered in step 1 of a regression predicting the associated outcome (e.g., job performance); then, the tag-contextualized version of the measure was entered in step 2. Results are depicted in Table 8. Tag-contextualized measures demonstrated incremental validity over instruction-contextualized measures for extraversion, openness to experience, and neuroticism.

Finally, relative weight analyses were conducted, which included only the instruction-contextualized and tag-contextualized versions of the predictors and directly compared their relative contributions toward predicting associated criteria. Tag-contextualized measures accounted for significantly more variance in the criteria in three cases (i.e., extraversion, openness to experience, and neuroticism). Results are summarized in Table 9. In no instance did instruction-level contextualization account for significantly more variance in the criteria than tag-level contextualization. This was true in both these analyses and those including all three predictors (i.e., base, instruction-contextualized, and

Table 5 Comparisons of base and work-contextualized personality's relationships with hypothesized criteria in primary study

Predictor	Criterion	Base <i>r</i>	Ins <i>r</i>	Tag <i>r</i>	Ins—Base <i>z</i>	Tag—Base <i>z</i>	Tag—Ins <i>z</i>
Extraversion	Job satisfaction	.31***	.29***	.38***	-0.64	2.13*	2.54*
Agreeableness	Perpetrated incivility	-.32***	-.35***	-.33***	-0.88	-0.19	0.68
Conscientiousness	Job performance	.17***	.17***	.18***	-0.12	0.21	0.33
Openness to Experience	Creative job performance	.30***	.35***	.45***	1.78*	4.84***	3.66***
Neuroticism	Emotional exhaustion	.52***	.52***	.64***	0.18	4.11***	4.15***

Base *r*, Ins (instruction) *r*, and Tag *r* denote the correlation between each version of the personality measure and its associated criterion. *N* = 399
 * *p* < .05; ** *p* < .01; *** *p* < .001

Table 6 Hierarchical regression analyses examining the incremental validity of work-contextualized personality with hypothesized criteria in primary study

Predictor	Ordered predictors	Criterion	<i>B</i>	ΔR^2	<i>R</i> ²
Extraversion	1. Base	Job satisfaction	0.21*	.09***	.09***
	2. Instruction		0.10	.00	.10***
Agreeableness	1. Base	Perpetrated incivility	-0.10	.11***	.11***
	2. Instruction		-0.20***	.02***	.13***
Conscientiousness	1. Base	Job performance	0.29	.03***	.03***
	2. Instruction		0.24	.00	.03***
Openness to experience	1. Base	Creative job performance	0.03	.09***	.09***
	2. Instruction		0.41***	.03***	.12***
Neuroticism	1. Base	Emotional exhaustion	0.32*	.27***	.27***
	2. Instruction		0.36***	.02***	.29***
	1. Base	Emotional exhaustion	0.14*	.27***	.27***
	2. Tag		0.72***	.15***	.42***

B values reflect the unstandardized coefficient from the final step of analysis; in all cases, the base personality measure was a significant predictor of the outcome before the contextualized measure was added. *N* = 399

* *p* < .05; ** *p* < .01; *** *p* < .001

tag-contextualized personality). Overall, these analyses provide support for hypotheses 2a, 2b, and 2c.

Supplemental Analyses

Although the focus of this work was on the effect of contextualization in established predictor-criterion relationships, examining all of the relationships in the data can provide a more comprehensive understanding of contextualization. Thus, as supplemental analyses, we examined the effects of contextualization on the non-hypothesized relationships (e.g., extraversion predicting perpetrated incivility, job performance, creative job performance, and emotional

exhaustion) using the same methods as above. All tables related to these supplemental analyses (i.e., Table S3–S7) are available in Appendix C.

Correlations between base personality, contextualized personality, and the non-hypothesized outcomes were compared using Steiger's (1980) *z*-tests; see Table S3. There were no significant differences between the base and instruction-contextualized versions of personality. Out of the 20 total comparisons, tag-contextualized personality demonstrated a stronger association with criteria than base personality in ten cases. Hierarchical regression analyses are presented in Table S4. The addition of the instruction-contextualized personality measure significantly improved

Table 7 Relative weight analyses assessing hypothesis 1

	R^2	Raw relative weights	95% CI ^a	Rescaled relative weights ^b	Predictor comparison CI ^c
Job satisfaction	.15				
Base extraversion		.04	[.01, .06]*	25.66	
Instruction extraversion		.03	[.01, .05]*	20.79	[-.03, .01]
Tag extraversion		.08	[.03, .13]*	53.56	[.003, .085]*
Perpetrated incivility	.13				
Base agreeableness		.04	[.01, .07]*	27.84	
Instruction agreeableness		.05	[.02, .09]*	39.30	[-.02, .05]
Tag agreeableness		.04	[.02, .08]*	32.87	[-.02, .04]
Job performance	.04				
Base conscientiousness		.01	[-.02, .03]	32.16	
Instruction conscientiousness		.01	[-.03, .03]	27.26	[-.02, .01]
Tag conscientiousness		.02	[-.02, .04]	40.58	[-.02, .03]
Creative job performance	.21				
Base openness to experience		.03	[.01, .05]*	15.61	
Instruction openness to experience		.05	[.02, .07]*	22.41	[-.003, .036]
Tag openness to experience		.13	[.08, .19]*	61.99	[.06, .15]*
Emotional exhaustion	.41				
Base neuroticism		.09	[.06, .13]*	22.60	
Instruction neuroticism		.09	[.06, .13]*	22.69	[-.04, .04]
Tag neuroticism		.23	[.17, .29]*	54.71	[.07, .21]*

^a95% confidence interval testing the statistical significance of the relative weights using a 10,000-iteration bootstrap

^bRescaled weights indicate the percentage of the explained variance that a predictor is responsible for

^cThese confidence intervals indicate whether the predictors significantly differ from the base predictor

*The 95% CI does not include 0

Table 8 Hierarchical regression analyses examining the incremental validity of tag-contextualized personality with hypothesized criteria in primary study

Predictor	Ordered predictors	Criterion	B	ΔR^2	R^2
Extraversion	1. Instruction	Job satisfaction	0.02	.09***	.09***
	2. Tag		0.37***	.06***	.14***
Agreeableness	1. Instruction	Perpetrated incivility	-0.19*	.12***	.12***
	2. Tag		-0.11	.01	.13*
Conscientiousness	1. Instruction	Job performance	0.21	.03***	.03***
	2. Tag		0.38	.01	.04***
Openness to experience	1. Instruction	Creative job performance	-0.06	.12***	.12***
	2. Tag		0.61***	.08***	.20***
Neuroticism	1. Instruction	Emotional exhaustion	0.13	.27***	.27***
	2. Tag		0.74***	.14***	.41***

B values reflect the unstandardized coefficient from the final step of analysis; in all cases, the instruction-contextualized personality measure was a significant predictor of the outcome before the tag-contextualized measure was added. $N=399$

* $p < .05$; ** $p < .01$; *** $p < .001$

prediction of outcomes over the base personality measure in nine cases out of 20. Tag-level contextualization demonstrated incremental validity over base personality in 14 of 20 cases. RWAs comparing base, instruction-contextualized, and tag-contextualized personality measures for

the non-hypothesized outcomes are presented in Table S5. Across 20 comparisons, instruction-contextualized personality never predicted significantly more variance in an outcome than base personality. The tag-contextualized version of a scale predicted significantly more variance in criteria in

Table 9 Relative weight analyses assessing hypothesis 2

	R^2	Raw relative weights	95% CI ^a	Rescaled relative weights ^b	Predictor comparison CI ^c
Job satisfaction	.14				
Instruction extraversion		.05	[.02, .08]*	31.50	
Tag extraversion		.10	[.05, .16]*	68.50	[.01, .10]*
Perpetrated incivility	.13				
Instruction agreeableness		.07	[.03, .12]*	54.41	
Tag agreeableness		.06	[.02, .11]*	45.59	[-.06, .03]
Job performance	.04				
Instruction conscientiousness		.02	[.0004, .0484]*	43.08	
Tag conscientiousness		.02	[.002, .056]*	56.92	[-.01, .03]
Creative job performance	.21				
Instruction openness to experience		.06	[.02, .09]*	28.34	
Tag openness to experience		.15	[.08, .21]*	71.66	[.04, .14]*
Emotional exhaustion	.41				
Instruction neuroticism		.14	[.09, .19]*	33.08	
Tag neuroticism		.28	[.20, .34]*	66.92	[.06, .23]*

^a95% confidence interval testing the statistical significance of the relative weights using a 10,000-iteration bootstrap

^bRescaled weights indicate the percentage of the explained variance that a predictor is responsible for

^cThese confidence intervals indicate whether the tag predictor significantly differs from the instruction predictor

*The 95% CI does not include 0

eight of 20 cases. Although the primary analyses focused on effects of contextualization in established predictor-criterion relationships, these supplemental analyses largely mirrored the results related to hypotheses 1a, 1b, and 1c. With the exception of conscientiousness, contextualization generally improved the relationships between personality and these work-related outcomes.

Turning to direct comparison of instruction- and tag-contextualized measures among the non-hypothesized outcomes, tag-contextualized personality evidenced significantly stronger correlations with outcomes than instruction-contextualized personality in six cases (see Table S3). In only one comparison did instruction-contextualized personality have a significantly stronger correlation with an outcome than tag-contextualized personality. Hierarchical regression analyses comparing instruction- and tag-contextualization among non-hypothesized outcomes are presented in Table S6. The addition of the tag-contextualized personality measure significantly improved prediction of the outcome over the instruction-contextualized personality measure in 14 of 20 cases. Table S7 displays the RWAs comparing instruction- and tag-level contextualization among non-hypothesized outcomes. Instruction-contextualized personality never predicted significantly more variance in an outcome than tag-contextualized personality. The tag-contextualized version of a scale predicted significantly more variance in outcomes in five of 20 cases. Although the

primary analyses for hypotheses 2a, 2b, and 2c focused on the effects of tag- and instruction-contextualization in established predictor-criterion relationships, these supplemental analyses also support the superiority of tag-level contextualization over instruction-level contextualization.

Discussion

Based on these findings, we are better able to understand how various personality factors are differentially impacted by contextualization. Specifically, we assessed the five factor model of personality and found that contextualization generally improves the strength of relationships and prediction between personality factors and work-relevant outcomes. These results replicate previous findings highlighting the benefits of utilizing contextualized measures to improve context-relevant prediction (Shaffer & Postlethwaite, 2012). Further, we examined the utility of the two most common and accessible forms of contextualization by comparing instruction-level and tag-level contextualizations in a within-person, multi-wave design. Our findings suggest tag-level contextualization outperformed instruction-level contextualization for extraversion, neuroticism, and openness to experience. Based on the pilot and primary study, we provide an explicit recommendation that if researchers and practitioners wish to contextualize measures, they should

prioritize the use of tags. More specifically, initial evidence supports implementing tag-level contextualization by altering between starting and ending items with tags. Seeing that the addition of instruction contextualization would have virtually no cost, researchers and practitioners may wish to employ contextualized instructions in addition to tags, assuming that this addition would not make for overly long or grammatically confusing instructions.

Theoretical Implications

The current study provides a methodologically rigorous replication and extension of previous findings demonstrating the utility of contextualization. The direct comparison of instruction-level and tag-level contextualization further advances the literature. Additionally, this study provides evidence for the inclusion of less frequently represented factors of personality (i.e., extraversion, openness to experience, neuroticism) in contextualization research. Especially in selection contexts, conscientiousness has been the dominant personality predictor (Barrick & Mount, 1991). Supporting the tenets of Cognitive-Affective Personality System theory, the current findings demonstrate the value of contextualizing multiple personality factors in order to predict a wide variety of valued workplace outcomes (Mischel & Shoda, 1995). Openness is often singled out as the personality trait that is the least work-relevant (Barrick & Mount, 1991). The current study demonstrated that openness predicts work-related outcomes better than generally assumed when contextualization is used. Also, our research utilizes more comprehensive and advanced analytical approaches to assess the utility of contextualized personality measures by directly comparing their relative contributions with relative weight analysis.

Based on the primary study's results, contextualization seemed to not significantly improve the relationship between conscientiousness and self-report job performance. Further, the supplemental analyses demonstrated that contextualization did not improve the relationship between conscientiousness and other work-related outcomes. One potential explanation that has been considered in previous research is the idea that the layman's interpretation of general conscientiousness may be heavily overlapped with the layman's representation of work (Shaffer & Postlethwaite, 2012). This finding has been found empirically in previous research as well (Heller et al., 2009). This conceptual overlap, in combination with Trait Activation Theory (Tett & Burnett, 2003) and Cognitive-Affective Personality System theory (Mischel & Shoda, 1995), suggests that individuals are already inclined to think of the work context when reading general conscientiousness items because the work context activates this trait. However, contextualization did significantly improve the relationship between conscientiousness and GPA in the pilot study and a markedly lower mean conscientiousness

score was observed in the base, non-contextualized group of the pilot study. This could be the result of differences between a student sample and a working sample. However, the unique timing of the pilot study data collection at the beginning of the COVID-19 pandemic suggests that, under certain conditions, unintended contexts may become particularly salient (Ansell et al., 2010; Ng et al., 2021). In other words, participants responding to the base conscientiousness items may have self-imposed a "pandemic" context. Thus, even if a particular personality factor is generally associated with a specific context (i.e., conscientiousness may be generally associated with work and/or school), contextualization may still prove beneficial to ensure this connection is made.

On the other hand, contextualization consistently improved the relationships and predictive ability of openness to experience, which is often viewed as the least work-relevant of the big five personality traits. Taken together, these findings suggest that some constructs (e.g., conscientiousness) may be more strongly, inherently associated with certain contexts (e.g., work) than other constructs (e.g., openness to experience). Along these lines, recent research has begun to explore the characteristics of items that may influence contextualization with the inception of "hidden framings" or implicit frames of reference that originate from item word choice or situational context (Schulze et al., 2021). Future research could help inform which personality traits or other constructs are more likely to benefit from contextualization. In line with Cognitive-Affective Personality System theory and Trait Activation Theory, the current contextualization research emphasizes the importance of psychological characteristics of current situations in impacting behavior (Mischel & Shoda, 1995; Tett & Burnett, 2003). However, Cognitive-Affective Personality System theory also posits that genes and early developmental history play important roles in determining behavior (Mischel & Shoda, 1995). Although less directly related to contextualization, future research should investigate the behavioral impacts of other predictors described by this theory.

Practical Implications

The current research provides actionable best practices for researchers and practitioners alike. First, we provide additional evidence that simple forms of contextualization do improve the predictive ability of common personality assessments. Practically speaking, there is little to no cost associated with contextualizing personality measures through adding tags or altering instructions. The benefits, however, can be significant. This is in contrast to complete contextualization, which may provide strong results but also requires extensive time and resources. Next, we provide initial evidence that tag-level contextualization should generally be preferred over instruction-level contextualization. Finally,

although future research should seek to replicate these findings, the results of our pilot study support the use of an alternating approach when applying tags to contextualize items.

As has been discussed in the extant literature, contextualized personality provides a potent predictor of important work and academic outcomes. At the same time, contextualized personality also provides better face validity for applicants when compared to general personality measures (Holtrop et al., 2014; Robie et al., 2017). Although we recommend the use of contextualized personality measures, one caveat should be considered. Generally speaking, the preponderance of evidence suggests that the best practice is to contextualize personality measures when the criteria of interest relate to the specific frame of reference (e.g., school, work). Thus, the bandwidth-fidelity dilemma should be considered in the decision to use contextualized personality measures. In other words, one should be sure to align the specificity of one's predictors with the specificity of one's outcomes.

Limitations and Future Directions

Although this study provides meaningful contributions to the extant literature and best practices around contextualized personality, there are several limitations worth considering. As mentioned, one of the downsides to complete contextualization is the validation work necessary to confirm the content domain has not changed due to the modifications. However, that is not to say this problem does not exist in less intrusive forms of contextualization. In fact, previous researchers have questioned whether tag-level contextualization impacts the psychometric properties of a scale (e.g., Robie & Risavy, 2016), although the current study showed support for the psychometric properties of the tag-level contextualized personality measures. In sum, although we endorse the utilization of tag-level contextualization, we also encourage researchers and practitioners to utilize these methods responsibly. We echo Heggstad and colleagues' (2019) recommendation that authors explicitly describe any changes they make to adjust a scale's context. Whenever possible, authors should provide their full list of contextualized items and/or contextualized instructions in an appendix or through an online supplement (e.g., housed on the Open Science Framework).

Next, we leveraged relatively stringent requirements to ensure high-quality data from our MTurk sample. These criteria may have had the effect of screening out less conscientious potential participants. Indeed, our participants reported relatively high average conscientiousness compared to samples in past research utilizing the mini-IPIP (Baldasaro et al., 2013; Donnellan et al., 2006). Thus, our findings related to conscientiousness may be partially attributable to an unusually conscientious sample. While our results support the

use of tag-level contextualization in research, the generalizability of our findings from a voluntary research study to a high-stakes selection scenario is an open question. To the extent that tags may be more effective than instruction-level contextualization because survey participants pay less attention to instructions, job applicants would likely be more motivated to carefully read instructions; thus, instruction-level contextualization may perform just as well as tags in such situations. Although previous research has supported the resilience of tag-level contextualization in hypothetical high-stakes scenarios, future research should examine whether our findings hold in selection settings (Bing et al., 2004; Schmit et al., 1995).

The last major limitation of the current study is the reliance on self-report data. Previous research has investigated potential common method variance effects that can impact observed effect sizes (Spector & Brannick, 2010). One key issue when considering common method variance is assessing whether the method in question (self-report) was chosen with rationale. Previous research has supported the accuracy of self-report college GPA as used in our pilot study (Caskie et al., 2014; Cassady, 2000) and job performance as used in the primary study (Williams & Levy, 1992), although other researchers have raised justifiable concerns about self-report job performance (e.g., Donaldson & Grant-Vallone, 2002). Aside from empirical support, there is also the question of whether the constructs conceptually should be assessed through a particular method. We argue that self-report assessment is clearly most appropriate for personality, emotional exhaustion, and job satisfaction. However, other forms of assessment may be more appropriate for job performance and incivility. Ultimately, future research should seek to replicate our findings with other operationalizations, such as supervisor-reported job performance or coworker-reported incivility.

In addition, new questions are worth considering regarding the way in which tags are implemented. In our pilot, the split tag condition outperformed the non-contextualized condition, while neither the primacy nor recency tag conditions significantly differed from the non-contextualized condition. As this is the first known research that compared placements of contextualized tags, future research should seek to replicate this finding. Further, although we provide possible explanations for the superiority of split tags, these explanations remain untested. We expect certain items may be clearer with a tag appended to the beginning as opposed to the end, or vice-versa. Future studies could investigate our grammar hypothesis by conducting think-aloud protocols with participants in order to assess whether there is noticeable conscious cognitive load associated with grammatically unusual or incorrect items (See Charters, 2003 for an introduction to think-aloud protocols). Alternatively, our unconscious scanning explanation could be tested using

eye-tracking software and tracking participant attention when presented with multiple items using repetitive tags.

Conclusion

This paper provides an in-depth investigation of accessible modifications practitioners and researchers can utilize to contextualize personality measures to evoke desired frames of reference. Specifically, instruction-level and tag-level contextualizations are directly compared in a within-person, time-lagged design. We provide best practice recommendations for the contextualization of personality measures. Specifically, the results generally support the use of tag-level contextualization and suggest that tags should be added to items in an alternating manner.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10869-024-09953-8>.

Acknowledgements The authors wish to thank Dr. Nathan Bowling for his helpful comments on an earlier version of this article.

Data Availability In accordance with the research participants' informed consent and, as approved by the relevant Institutional Review Board, data associated with this manuscript are not publicly available.

Declarations

Conflict of Interest The authors declare no competing interests.

References

- Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations. *Journal of Management*, 47(4), 823–837. <https://doi.org/10.1177/0149206320969787>
- Ajzen, I. (2005). Laws of human behavior: Symmetry, compatibility, and attitude-behavior correspondence. In A. Beauducel, B. Biehl, M. Bosniak, W. Conrad, G. Schonberger, & D. Wagener (Eds.), *Multivariate research strategies* (pp. 3–19). Shaker.
- Ansell, C., Boin, A., & Keller, A. (2010). Managing transboundary crises: Identifying the building blocks of an effective response system. *Journal of Contingencies and Crisis Management*, 18(4), 195–207.
- Baldasaro, R. E., Shanahan, M. J., & Bauer, D. J. (2013). Psychometric properties of the mini-IPIP in a large, nationally representative sample of young adults. *Journal of Personality Assessment*, 95(1), 74–84.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1), 1–26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>
- Bing, M. N., Whanger, J. C., Davison, 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. <https://doi.org/10.1037/0021-9010.89.1.150>
- Bing, M. N., Davison, H. K., & Smothers, J. (2014). Item-level frame-of-reference effects in personality testing: An investigation of incremental validity in an organizational setting. *International Journal of Selection and Assessment*, 22(2), 165–178. <https://doi.org/10.1111/ijsa.12066>
- Bowling, N. A., & Burns, G. N. (2010). A comparison of work-specific and general personality measures as predictors of work and non-work criteria. *Personality and Individual Differences*, 49(2), 95–101. <https://doi.org/10.1016/j.paid.2010.03.009>
- Bowling, N. A., & Hammond, G. D. (2008). A meta-analytic examination of the construct validity of the Michigan Organizational Assessment Questionnaire Job Satisfaction Subscale. *Journal of Vocational Behavior*, 73(1), 63–77. <https://doi.org/10.1016/j.jvb.2008.01.004>
- Cammann, C., Fichman, M., Jenkins, D., & Klesh, J. (1983). Assessing the attitudes and perceptions of organizational members. In S. E. Seashore, E. E. Lawler, P. H. Mirvis, & C. Cammann (Eds.), *Assessing organizational change: A guide to methods, measures, and practices* (pp. 71–138). Wiley.
- Caskie, G. I., Sutton, M. C., & Eckhardt, A. G. (2014). Accuracy of self-reported college GPA: Gender-moderated differences by achievement level and academic self-efficacy. *Journal of College Student Development*, 55(4), 385–390. <https://doi.org/10.1353/csd.2014.0038>
- Cassady, J. C. (2000). Self-reported GPA and SAT: A methodological note. *Practical Assessment, Research, and Evaluation*, 7(12), 1–4. <https://doi.org/10.7275/5hym-y754>
- Charters, E. (2003). The use of think-aloud methods in qualitative research. An introduction to think-aloud methods. *Brook Education Journal*, 12(2), 68–82. <https://doi.org/10.26522/brocked.v12i2.38>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). Applied multiple regression/correlation analysis for the behavioral sciences. Routledge. <https://doi.org/10.4324/9780203774441>
- Cortina, L. M., Magley, V. J., Williams, J. H., & Langhout, R. D. (2001). Incivility in the workplace: Incidence and impact. *Journal of Occupational Health Psychology*, 6(1), 64–80. <https://doi.org/10.1037/1076-8998.6.1.64>
- Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, 17, 245–260.
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, 18(2), 192–203. <https://doi.org/10.1037/1040-3590.18.2.192>
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, 91(1), 40–57. <https://doi.org/10.1037/0021-9010.91.1.40>
- Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden Lehrbuch*. Beltz.
- Furnham, A., Monsen, J., & Ahmetoglu, G. (2009). Typical intellectual engagement, big five personality traits, approaches to learning and cognitive ability predictors of academic performance. *British Journal of Educational Psychology*, 79(4), 769–782. <https://doi.org/10.1348/978185409X412147>
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1), 26–42. <https://doi.org/10.1037/1040-3590.4.1.26>
- Gray, E. K., & Watson, D. (2002). General and specific traits of personality and their relation to sleep and academic performance. *Journal of Personality*, 70(2), 177–206. <https://doi.org/10.1111/1467-6494.05002>
- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18(2), 135–164. <https://doi.org/10.1111/j.1744-6570.1965.tb00273.x>
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and

- meta-analysis. *Personnel Psychology*, 57(3), 639–683. <https://doi.org/10.1111/j.1744-6570.2004.00003.x>
- Heggstad, E. D., Scheaf, D. J., Banks, G. C., Monroe Hausfeld, M., Tonidandel, S., & Williams, E. B. (2019). Scale adaptation in organizational science research: A review and best-practice recommendations. *Journal of Management*, 45(6), 2596–2627. <https://doi.org/10.1177/0149206319850280>
- 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(4), 1051–1084.
- Holbrook, A. L., Krosnick, J. A., Moore, D., & Tourangeau, R. (2007). Response order effects in dichotomous categorical questions presented orally: The impact of question and respondent attributes. *The Public Opinion Quarterly*, 71(3), 325–348. <https://doi.org/10.1093/poq/nfm024>
- Holtrop, D., Born, M. P., de Vries, A., & de Vries, R. E. (2014). A matter of context: A comparison of two types of contextualized personality measures. *Personality and Individual Differences*, 68, 234–240. <https://doi.org/10.1016/j.paid.2014.04.029>
- Holtz, B. C., Ployhart, R. E., & Dominguez, A. (2005). Testing the rules of justice: The effects of frame-of-reference and pre-test validity information on personality test responses and test perceptions. *International Journal of Selection & Assessment*, 13(1), 75–86. <https://doi.org/10.1111/j.0965-075X.2005.00301>
- Hunthausen, J. M., Truxillo, D. M., Bauer, T. N., & Hammer, L. B. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology*, 88(3), 545–551. <https://doi.org/10.1037/0021-9010.88.3.545>
- Johnson, R. E., Rosen, C. C., & Djurdjevic, E. (2011). Assessing the impact of common method variance on higher order multidimensional constructs. *Journal of Applied Psychology*, 96(4), 744–761. <https://doi.org/10.1037/a0021504>
- Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, 87(3), 530–541. <https://doi.org/10.1037/0021-9010.87.3.530>
- Kammeyer-Mueller, J. D., Simon, L. S., & Judge, T. A. (2016). A head start or a step behind? Understanding how dispositional and motivational resources influence emotional exhaustion. *Journal of Management*, 42(3), 561–581. <https://doi.org/10.1177/0149206313484518>
- Kelley, M. R., Neath, I., & Surprenant, A. M. (2013). Three more semantic serial position functions and a SIMPLE explanation. *Memory & Cognition*, 41(4), 600–610. <https://doi.org/10.3758/s13421-012-0286-1>
- Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51(2), 201–219. <https://doi.org/10.1086/269029>
- 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. <https://doi.org/10.1037/0021-9010.93.2.268>
- Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Organizational Behavior*, 2(2), 99–113. <https://doi.org/10.1002/job.4030020205>
- McAbee, S. T., & Oswald, F. L. (2013). The criterion-related validity of personality measures for predicting GPA: A meta-analytic validity competition. *Psychological Assessment*, 25(2), 532–544. <https://doi.org/10.1037/a0031748>
- Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 80, 252–283. <https://doi.org/10.1037/h0035002>
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, 102(2), 246–268. <https://doi.org/10.1037/0033-295X.102.2.246>
- Ng, M. A., Naranjo, A., Schlotzhauer, A. E., Shoss, M. K., Kartvelishvili, N., Bartek, M., ... & Silva, C. (2021). Has the COVID-19 pandemic accelerated the future of work or changed its course? Implications for research and practice. *International Journal of Environmental Research and Public Health*, 18(19), 10199.
- 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*, 93(1), 116–130. <https://doi.org/10.1037/0022-3514.93.1.116>
- Oldham, G. R., & Cummings, A. (1996). Employee creativity: Personal and contextual factors at work. *Academy of Management Journal*, 39(3), 607–634. <https://doi.org/10.2307/256657>
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867–872. <https://doi.org/10.1016/j.jesp.2009.03.009>
- Pace, V. L., & Brannick, M. T. (2010). Improving prediction of work performance through frame-of-reference consistency: Empirical evidence using openness to experience. *International Journal of Selection & Assessment*, 18(2), 230–235. <https://doi.org/10.1111/j.1468-2389.2010.00506.x>
- Pathki, C. S., Kluemper, D. H., Meuser, J. D., & McLarty, B. D. (2022). The org-B5: Development of a short work frame-of-reference measure of the big five. *Journal of Management*, 48(5), 1299–1337. <https://doi.org/10.1177/01492063211002627>
- Pearce, J. L., & Porter, L. W. (1986). Employee responses to formal performance appraisal feedback. *Journal of Applied Psychology*, 71(2), 211–218. <https://doi.org/10.1037/0021-9010.71.2.211>
- Ployhart, R. E., Ziegert, J. C., & McFarland, L. A. (2003). Understanding racial differences on cognitive ability tests in selection contexts: An integration of stereotype threat and applicant reactions research. *Human Performance*, 16(3), 231–259. https://doi.org/10.1207/S15327043HUP1603_4
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539–569.
- 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. <https://doi.org/10.1111/j.1468-2389.2011.00540.x>
- Robie, C., & Risavy, S. D. (2016). A comparison of frame-of-reference and frequency-based personality measurement. *Personality and Individual Differences*, 92, 16–21. <https://doi.org/10.1016/j.paid.2015.12.005>
- 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(4), 348–365. <https://doi.org/10.1177/109442810034003>
- Robie, C., Born, M. P., & Schmit, M. J. (2001). Personal and situational determinants of personality responses: A partial reanalysis and reinterpretation of the Schmit et al. (1995) data. *Journal of Business and Psychology*, 16(1), 101–117. <https://doi.org/10.1023/A:1007843906550>
- Robie, C., Risavy, S. D., Holtrop, D., & Born, M. P. (2017). Fully contextualized, frequency-based personality measurement: A replication and extension. *Journal of Research in Personality*, 70, 56–65. <https://doi.org/10.1016/j.jrp.2017.05.005>

- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2021). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology, 107*(11), 2040–2068. <https://doi.org/10.1037/apl0000994>
- Schilpzand, M. C., Herold, D. M., & Shalley, C. E. (2011). Members' openness to experience and teams' creative performance. *Small Group Research, 42*(1), 55–76. <https://doi.org/10.1177/1046496410377509>
- 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*(5), 607–620. <https://doi.org/10.1037/0021-9010.80.5.607>
- Schulz, J., West, S. G., Freudenstein, J. P., Schäpers, P., Mussel, P., Eid, M., & Krumm, S. (2021). Hidden framings and hidden asymmetries in the measurement of personality—A combined lens-model and frame-of-reference perspective. *Journal of Personality, 89*(2), 357–375. <https://doi.org/10.1111/jopy.12586>
- Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65*(3), 445–494. <https://doi.org/10.1111/j.1744-6570.2012.01250.x>
- Shamon, H., & Berning, C. (2020). Attention check items and instructions in online surveys with incentivized and non-incentivized samples: Boon or bane for data quality? *Survey Research Methods, 14*(1), 55–77. <https://doi.org/10.2139/ssrn.3549789>
- Sosnowska, J., De Fruyt, F., & Hofmans, J. (2019). Relating neuroticism to emotional exhaustion: A dynamic approach to personality. *Frontiers in Psychology, 10*, 2264. <https://doi.org/10.3389/fpsyg.2019.02264>
- Spector, P. E., & Brannick, M. T. (2010). Common method issues: An introduction to the feature topic in organizational research methods. *Organizational Research Methods, 13*(3), 403–406. <https://doi.org/10.1177/1094428110366303>
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*(2), 245–251. <https://doi.org/10.1037/0033-2909.87.2.245>
- Swift, V., & Peterson, J. B. (2019). Contextualization as a means to improve the predictive validity of personality models. *Personality and Individual Differences, 144*, 153–163. <https://doi.org/10.1016/j.paid.2019.03.007>
- Taylor, S. G., & Kluepfer, D. H. (2012). Linking perceptions of role stress and incivility to workplace aggression: The moderating role of personality. *Journal of Occupational Health Psychology, 17*(3), 316–329. <https://doi.org/10.1037/a0028211>
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*(3), 500–517. <https://doi.org/10.1037/0021-9010.88.3.500>
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*(4), 703–742. <https://doi.org/10.1111/j.1744-6570.1991.tb00696.x>
- Tonidandel, S., & LeBreton, J. M. (2015). RWA web: A free, comprehensive, web-based, and user-friendly tool for relative weight analyses. *Journal of Business and Psychology, 30*(2), 207–216. <https://doi.org/10.1007/s10869-014-9351-z>
- Tonidandel, S., LeBreton, J. M., & Johnson, J. W. (2009). Determining the statistical significance of relative weights. *Psychological Methods, 14*(4), 387–399. <https://doi.org/10.1037/a0017735>
- Welbourne, J. L., Miranda, G., & Gangadharan, A. (2020). Effects of employee personality on the relationships between experienced incivility, emotional exhaustion, and perpetrated incivility. *International Journal of Stress Management, 27*(4), 335–345. <https://doi.org/10.1037/str0000160>
- Williams, J. R., & Levy, P. E. (1992). The effects of perceived system knowledge on the agreement between self-ratings and supervisor ratings. *Personnel Psychology, 45*(4), 835–847. <https://doi.org/10.1111/j.1744-6570.1992.tb00970.x>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.