SCANDINAVIAN JOURNAL OF WORK AND ORGANIZATIONAL PSYCHOLOGY

Can We Improve the Assessment of Work Characteristics by Completely Contextualizing Items?

RESEARCH NOTE

BERIT GREULICH
CORNELIUS J. KÖNIG
CARLA V. MEIXENSBERGER

*Author affiliations can be found in the back matter of this article

STOCKHOLM UNIVERSITY PRESS

ABSTRACT

Employee surveys are a cost-effective way to assess work-related stress, but their generic formulation often allows for broad interpretation. Research from personality assessment suggests that contextualizing survey items may enhance reliability and validity by narrowing this interpretive scope, especially when items are fully adapted to fit a specific context ("complete contextualization"). This study investigates whether these benefits also apply to assessing work characteristics, using a sample of 323 German medical assistants. Items from the Copenhagen Psychosocial Questionnaire were tailored to the medical practice context with expert input. Participants completed both generic and contextualized versions of the items in a counterbalanced order, with psychological strain as the criterion. The results showed some positive effects, as the contextualized versions partly explained more criterion variance. However, contextualization did not enhance reliability or participants' reactions and led to longer items. Overall, complete contextualization may improve work characteristics assessments; however, further research is needed to evaluate its broader application.

CORRESPONDING AUTHOR:
Cornelius J. König
Saarland University, Campus A1 3, 66123 Saarbrücken, Germany
cornelius.koenig@uni-saarland.de

KEYWORDS:
work-related stress; frame-of-reference; work stress surveys; reliability; validity

TO CITE THIS ARTICLE:
Greulich, B., König, C. J., & Meixensberger, C. V. (2025). Can We Improve the Assessment of Work Characteristics by Completely Contextualizing Items? *Scandinavian Journal of Work and Organizational Psychology*, 10(1): 13, 1–10. DOI: https://doi.org/10.16993/sjwop.348

To counteract the negative effects of work stress at an early stage, managers need to know the extent of stressors and resources in their organization so that they can successfully implement interventions (Havermans et al. 2018). A common practice for assessing work characteristics as potential stressors is through self-reports in employee surveys, which usually ask about the frequency and intensity of different characteristics (e.g., time pressure, work interruptions) and resources (e.g., leadership quality, communication possibilities) at the workplace. This method directly reflects employees' perspectives and has the advantage of being economical and simple to implement (Eatough & Spector 2013). Nonetheless, the scope for interpretation of the items poses a notable challenge: Employee survey items are generally formulated in a generic and non-contextualized manner, which can lead to variations in subjects' interpretations of the items, and this variability in interpretation needs to be addressed to ensure accurate and unbiased measurement of constructs (Greulich et al. 2021).

To counteract the challenge of generic items, researchers interested in assessing personality have advocated for adding relevant context to questionnaires (Holtz, Ployhart, & Dominguez 2005). This contextualization creates a consistent frame-of-reference that should have positive effects on criterion validity and reliability (Holtrop et al. 2014) due to the improvement of two quality criteria of questionnaires: increased within-person consistency due to responders rating items with a consistent frame-of-reference, and reduced between-person variability due to a high number of respondents using a consistent frame-of-reference (see also Ajzen & Fishbein 1977; Lievens, De Corte, & Schollaert 2008). The present study aims to test whether the positive effects of contextualization can be transferred from the personality field to work characteristics survey items.

HYPOTHESES DEVELOPMENT

Researchers have identified several work-related risk factors linked to increased rates of common mental health problems (e.g., low job control, high job demands, bullying, and role conflicts, Sonnentag, Tay, & Nesher Shoshan 2023). According to occupational health psychology models such as the Job Demands-Control model (Karasek 1979), the frequency and severity of job demands (e.g., time pressure) lead to increased effort, which depletes employees' physical, emotional, and cognitive resources, leading to exhaustion and stress among employees and, over time, to serious health problems. Job resources (e.g., task control) are assumed to serve as a kind of buffer and therefore weaken the impact of job demands on strain (Karasek 1979).

Due to the negative consequences of work-related stress, organizations have an interest in identifying the nature and extent of potential stressors and resources. This creates the basis for the development and implementation of measures to prevent negative outcomes. Work-related stressors and resources (i.e., work characteristics) are usually assessed using self-reports (e.g., Kristensen et al. 2005). These are surveys that can be completed by employees as part of an employee survey by rating stressors and resources in terms of their frequency and intensity. Typical characteristics measured with such questionnaires include time pressure, concentration demands, and one-sided physical demands; typical resources surveyed include task control, support from colleagues or supervisors, and development opportunities.

Self-reports are widely used for assessing work characteristics because they are convenient, cost-effective, and employees are typically more familiar with their work environment than external evaluators (Semmer, Zapf, & Dunckel 1995). Furthermore, stress research emphasizes that the subjective appraisal of stressors and resources plays a key role in the perception of stress (Lazarus & Folkman 1984). For such surveys to remain cost-effective, they need to be applicable across various workplaces. As a result, these questionnaires often contain generically phrased items, leaving ample room for individual interpretation. However, answering survey questions is a complex cognitive process vulnerable to biases (Greulich et al. 2021; Tourangeau 1987). Scholars argue that respondents may interpret generic items differently, imbuing each with personal context (e.g., Credé, Bashshur, & Niehorster 2010; Holtrop et al. 2014; Schulze et al. 2021). This wide scope for interpretation can impair measurement accuracy, as interindividual differences in interpretation may reduce criterion validity (Lievens, De Corte, & Schollaert 2008). Inaccurate measures, in turn, could hinder the identification of psychosocial factors that precede strain, ultimately leading to misguided implementation of management's preventive measures.

To reduce individual interpretation of personality scale items, researchers have proposed contextualizing survey items: embedding them within a relevant context (Schmit et al. 1995). Various types of contextualization exist, differing in their level of detail (e.g., Holtrop et al. 2014; Schlotzhauer, Ng, & Su 2025). At a low degree, contextualization occurs through an instruction in which participants are asked to think of a particular situation while completing the survey. Medium-level contextualization, known as "tagged contextualization," involves adding a tag to each item, indicating a specific environment or time period that respondents should consider when answering (e.g., Holtrop et al. 2014; Schlotzhauer, Ng, & Su 2025). The highest level of contextualization is achieved when items are completely

rewritten to reflect a particular context, known as "complete contextualization" (e.g., Holtrop et al. 2014).

If contextualizing survey items helps to limit individual interpretation, it should also enhance the internal consistency (i.e., reliability) of the measures. Several studies have indeed demonstrated that contextualizing personality survey items improves reliability (e.g., Lievens, De Corte, & Schollaert 2008; for a different finding, see Holtrop, Born, & de Vries 2014). Moreover, contextualization has been shown to enhance the criterion-related validity of personality scales compared to non-contextualized versions, likely due to reduced error variance (Sackett et al. 2022; Shaffer & Postlethwaite 2012).

Although contextualization of items has primarily been studied in the domain of personality assessment (e.g., Credé, Bashshur, & Niehorster 2010; Schmit et al. 1995; Schulze et al. 2021), contextualization might also be beneficial for assessing work characteristics, as responding to such items is influenced by uncertainty about item interpretation (Greulich et al. 2021). Indeed, it has already been shown that adding a time tag to a job insecurity scale can influence response behavior (Debus et al. 2019), because it reduces error in measurement. In addition, adding a social comparison tag to work stress items has been demonstrated to increase the criterion validity of these scales (Greulich, König, & Fischer 2024). However, the benefits of completely contextualizing work characteristics items remain unexplored, although contextualization should also increase measurement accuracy in the case of work characteristics. This is surprising because Holtrop et al. (2014) reported that complete contextualization of personality items increases criterion validity more effectively than tagged contextualization (although this was not found in Holtrop, Born, & de Vries 2014). Thus, this study aimed to test the effects of complete contextualization on reliability and criterion validity, with psychological strain as the criterion. The General Health Questionnaire (GHQ-12; Goldberg & Hillier 1979; Linden et al. 1996) was used as it is a widely established measure of psychological strain in occupational health psychology (e.g., Fryers et al. 2004), which has the added advantage of avoiding confounding health problems with their assumed work-related causes. More formally, we hypothesized:

Hypothesis 1a: Complete contextualized work characteristics items have a higher internal consistency than generic work characteristics items.

Hypothesis 1b: Complete contextualized work characteristics items have a higher criterion validity (i.e., with psychological strain) than generic work characteristics items.

Participants may also respond differently to contextualized items. According to Gilliland's (1993) model of applicant reactions, contextualized items should be preferred by participants as they likely resonate more closely with their personal experiences compared to generic items. In fact, Holtrop et al. (2014) found that contextualized items were associated with increased perceived predictive and face validity, although participants unexpectedly favored generic items in terms of overall preference. Since our participants were employees, contextualizing work characteristics items should enhance their understanding of their current work situation, potentially fostering a sense of fairness when responding (Gilliland, 1993). Following Holtrop et al. (2014), we examined three participant reactions: liking of the different item versions, face validity (i.e., perceived relevance), and perceived predictive validity (for their working performance). We expected that participants would react more positively to the completely contextualized work characteristics items. Thus, we hypothesized:

Hypothesis 2: Complete contextualized work characteristics items will achieve a higher approval rating for participants' reactions (liking, face validity, and perceived validity for predicting work performance) than generic work characteristics items.

METHODS

SAMPLE

The sample consists of 323 German medical assistants in doctor's offices of different specialties (e.g., family medicine, gynecology, cardiology, etc.). To recruit them, we contacted randomly selected doctors' offices in two towns, and additional data were collected in a training center for medical assistants. Informed consent was obtained from all participants. Nearly all participants were female (99.6%) and aged between 18 and 63 years (M = 36.49, SD = 11.62). On average, participants were employed for 15.17 years (SD = 11.28), and 66.5% were working full-time. Ethical review and approval were not required for this study in accordance with the local legislation and the requirements of Saarland University. The data is available at https://osf.io/8k6vw/?view_only=ca5ce19ae2ab4c8da13d0fa6a30f7689.

PROCEDURE

The study was conducted via a paper-and-pencil questionnaire, which consisted of demographic data, a generic version of work characteristics items and participants' reactions to them, a contextualized version of work characteristics items and participants' reactions to them, as well as a psychological strain scale. The order of the generic and contextualized versions was

counterbalanced. The questionnaires were distributed to doctors' offices by a research assistant. After completion of the questionnaire, the participants placed them in a ballot box that was picked up after approximately two weeks.

MEASURES

Work characteristics survey. With the assistance of a physician and a medical assistant, relevant stressor and resource scale items were selected from the German adaptation of the Copenhagen Psychosocial Questionnaire (COPSOQ; Nübling et al. 2005), with the COPSOQ being a widely used work stress survey in Germany, initially developed in Denmark by Kristiansen et al. (2005). The two experts identified only a subset of items as relevant to the everyday work of medical assistants and suitable for complete contextualization. In total, 13 items were selected and grouped into five scales. Three scales assessed stressors: quantitative demands, emotional demands, and hiding emotions. Two scales measured work resources: sense of community and quality of leadership. All items were rated on a five-point Likert scale ranging from 1 (*never*) to 5 (*always*).

For the generic version, the selected items remained unchanged. To create the fully contextualized version (see Table 1), we used Holtrop et al.'s procedure (2014). First, two research assistants independently developed contextualized versions of the selected generic items. They then discussed and agreed upon a finalized version for each item. Finally, a medical assistant attempted to match the contextualized items to their corresponding generic ones. If successful, the contextualized item was retained; if not, the process was repeated. The goal of contextualization was to establish a clear reference to the daily work of medical assistants by incorporating specific behavioral contexts and using job-specific terminology (e.g., "patient"). Correlations between stressor and resource scales in both the generic and contextualized versions (see Table 2) suggest that the underlying constructs remained fairly consistent despite the modifications.

Participants' reactions. Reactions were measured as in Holtrop et al. (2014). Directly after completing each version of the work stress survey, participants' reactions to the particular version (generic/contextualized) were measured via three items. For liking, an item based on Wiechmann and Ryan (2003) was used: *I did not enjoy completing this questionnaire* (reverse coded). For face validity and perceived predictive validity, items from Smither et al. (1993) were used: *The content of this questionnaire is clearly related to my work* and *With the results of this questionnaire my work performance can be predicted*. Participants reacted on a five-point Likert scale, ranging from 1 (*completely disagree*) to 5 (completely agree). Holtrop et al. (2014) analyzed on the item-level, and we followed their example.

SCALE	GENERIC VERSION	CONTEXTUALIZED VERSION
Quantitative demands	Do you have to work very fast?	In the everyday work of a medical assistant, there are situations in which many things have to be done as quickly as possible. For example, careless mistakes can happen when coordinating appointments due to time pressure. How often do you have to hurry when processing tasks?
	How often do you not have time to complete all your work tasks?	In the everyday work of a medical assistant, it can happen that a lot of tasks have to be handled. How often does it happen, for example, that you do not have enough time for administrative tasks because you also have to take care of the patients' needs on site, or vice versa?
	Do you have to do overtime?	In the everyday work of a medical assistant, it can happen that patients who do not have an appointment come to the practice shortly before closing time. This means that working hours are extended. How often does this happen?
Emotional demands	Do you have to deal with other people's personal problems as part of your work?	In the everyday work of a medical assistant, it is part of the job to be in contact with patients and relatives. How often do you have to deal with patients" personal problems?
	Is your work emotionally demanding?	Confronting the suffering of patients is part of the everyday work of a medical assistant. How often do you find that their fate is very close to you and places an emotional burden on you?
Hiding emotions	Does your work require that you hide your feelings?	The everyday work of a medical assistant involves many emotional situations. Feelings such as sadness or anger in contact with patients are not uncommon. How often do you have to suppress your feelings to appear "neutral" to others?
	Does your work require that you do not state your opinion?	The everyday work of a medical assistant involves many conflictual situations. Nevertheless, you must always maintain your professionalism. How often does it happen that you have to show an opinion to others in a situation that is not in line with your actual opinion?

(Contd.)

SCALE	GENERIC VERSION	CONTEXTUALIZED VERSION
Sense of community	Is there a good atmosphere between you and your colleagues?	In the everyday work of a medical assistant, teamwork plays a major role. Good teamwork has a positive effect on morale. How often do you feel that the cooperation between you and your colleagues is harmonious?
	Is there good co-operation between the colleagues at work?	In the everyday work of a medical assistant, patient care and administrative tasks should be understood as teamwork. How well does co-operation between colleagues work?
Quality of leadership	To what extent would you say that your immediate superior makes sure that the members of staff have good development opportunities?	In the everyday work of a medical assistant, it is important to constantly develop both social and professional skills. How often does your immediate supervisor attach importance to training and other opportunities for further development?
	To what extent would you say that your immediate superior gives high priority to job satisfaction?	Job satisfaction is very important in the everyday work of a medical assistant as it influences everyday contact with patients and colleagues, for example. How often does your immediate supervisor take time to talk to you about how you are feeling?
	To what extent would you say that your immediate superior is good at work planning?	Good planning and organization are essential to ensure that practice processes run smoothly. How well is your immediate supervisor structured at work, and does it allow for good planning of procedures in the doctor's office?
	To what extent would you say that your immediate superior is good at solving conflicts?	In the everyday work of a medical assistant, problems can arise due to conflicts in the team or too high a workload. How often does your immediate supervisor give you the opportunity to voice these problems and offer possible solutions?

Table 1 Generic and Contextualized Work Characteristics Items, Chosen from the German Adaptation of the Copenhagen Psychosocial Questionnaire (Nübling et al., 2005).

Note. All items are originally in German.

Psychological strain. To test the effect of complete contextualization on criterion validity, we used psychological strain as a criterion, using the GHQ-12 (Goldberg & Hillier 1979; Linden et al. 1996) as the measure. This 12-item version has a four-point response scale, with responses ranging from 0 (*not at all*) to 3 (*much more than usual*), resulting in a possible range of 0 to 36; higher scores indicate a higher level of psychological strain.

STATISTICAL METHODS

To analyze changes in the internal consistency (i.e., H1a), we compared them using the computer program Alpha Test, a program for testing hypotheses about Cronbach's alpha using chi-square (Lautenschlager & Meade 2008). It builds on the work of Feldt, Woodruff, & Salih (1987), who developed mathematical formulas for the comparison of two alphas while taking the dependency between the coefficients into account. To examine the criterion validity of the scales (i.e., H1b), we used hierarchical regression analyses to test how much one version of a scale explains variance above and beyond the other version. To assess differences in participants' reactions (i.e., H2), we used *t*-tests for paired samples.

RESULTS

For reliability analysis, Cronbach's alpha was calculated for each scale (reported in Table 2). The item analysis revealed that the fourth item of the contextualized scale "quality of leadership" had to be excluded from the generic and contextualized version due to insufficient discriminatory power ($r < 0.3$). Most of the generic stressor and resource scales used in this study showed acceptable to high reliability (0.65 to 0.88); the generic scale "quantitative demands" had a rather low Cronbach's alpha (0.59). The internal consistency of the stressor scales "quantitative demands" and "emotional demands" did not differ between the generic and contextualized versions, $\chi^2(1) = 1.51$, $p = 0.219$ (for "quantitative demands"), $\chi^2(1) = 2.10$, $p = 0.148$ (for "emotional demands"). All other scales decreased significantly in internal consistency due to item contextualization, "hiding emotions": $\chi^2(1) = 6.02$, $p = 0.014$; "sense of community": $\chi^2(1) = 12.12$, $p < 0.001$; "quality of leadership": $\chi^2(1) = 50.47$, $p < 0.001$. These results therefore did not support H1a.

Table 3 reports the results regarding H1b (i.e., the effects on criterion validity). Hierarchical regression analyses showed that the completely contextualized version of the scales explained additional variance in the GHQ beyond the generic version for the scales "quantitative demands," "hiding emotions," and "sense of community," as indicated by a significant increase in explained variance (i.e., ΔR^2). Such an increase in ΔR^2 was not found for the scales "emotional demands" and "quality of leadership." In an additional set of analyses, we examined whether the generic scales provided incremental validity over the contextualized versions (see Table 4). These analyses revealed that the generic

		M	*SD*	α	1	2	3	4	5	6	7	8	9	10
Predictor														
1	(Generic) Quantitative demands	3.38	0.69	0.59										
2	(Generic) Emotional demands	3.53	0.87	0.65	0.34**									
3	(Generic) Hiding emotions	3.03	1.01	0.73	0.18**	0.39**								
4	(Generic) Sense of community	4.13	0.75	0.88	−0.12*	−0.02	−0.10							
5	(Generic) Quality of leadership	3.30	0.99	0.85	−0.25**	−0.09	−0.18**	0.33**						
6	(Contextualized) Quantitative demands	3.48	0.72	0.65	0.62**	0.43**	0.23**	−0.12*	−0.19**					
7	(Contextualized) Emotional demands	3.11	0.77	0.56	0.21**	0.55**	0.23**	−0.03	0.04	0.33**				
8	(Contextualized) Hiding emotions	3.04	0.76	0.58	0.11*	0.26**	0.35**	−0.12	−0.13*	0.22**	0.32**			
9	(Contextualized) Sense of community	3.88	0.79	0.83	−0.16**	−0.04	−0.10	0.73**	0.37**	−0.15**	−0.01	−0.15**		
10	(Contextualized) Quality of leadership	3.21	0.90	0.69	−0.20**	−0.11*	−0.21**	0.32**	0.79**	−0.15*	0.04	−0.19**	0.33**	
Criterion														
11	GHQ score	2.08	0.53	0.90[a]	0.33**	0.16**	0.22**	−0.28**	−0.32**	0.32**	0.16**	0.25**	−0.34**	−0.29**

Table 2 Correlations for Predictor Variables and the Criterion Variable.

Note. $N = 323$. α = Cronbach's alpha. GHQ = General Health Questionnaire. [a]This reliability coefficient is a bit higher than meta-analytical estimate reported by Wojujutari, Idemudia, and Ugwu (2024).

*$p < 0.05$, ** $p < 0.01$.

Variable	Model	*R*	R^2	R^2_{adj}	*SE*	ΔR^2	F_{change}	df_1	df_2	*p*
Quantitative demands	1[a]	0.33	0.11	0.11	0.50	0.11	39.46	1	321	<0.001
	2[b]	0.36	0.13	0.13	0.50	0.02	7.54	1	320	0.006
Emotional demands	1[a]	0.16	0.03	0.02	0.53	0.03	8.63	1	320	0.004
	2[b]	0.19	0.03	0.03	0.53	0.01	2.63	1	319	0.150
Hiding demands	1[a]	0.22	0.05	0.05	0.52	0.05	16.97	1	321	<0.001
	2[b]	0.29	0.09	0.08	0.51	0.03	12.03	1	320	<0.001
Sense of community	1[a]	0.28	0.08	0.07	0.51	0.08	26.63	1	319	<0.001
	2[b]	0.34	0.12	0.11	0.50	0.04	13.51	1	318	<0.001
Quality of leadership	1[a]	0.32	0.10	0.10	0.51	0.10	35.78	1	321	<0.001
	2[b]	0.32	0.11	0.10	0.51	0.00	1.60	1	320	0.206

Table 3 Results of Hierarchical Regression Analyses for Psychological Strain: Variance Explained by the Contextualized Version Beyond the Generic Version.

Note. [a]Predictor: Generic version; [b]Predictors: Generic version and contextualized version. $N = 323$.

VARIABLE	MODEL	R	R^2	R^2_{adj}	SE	ΔR^2	F_{change}	df_1	df_2	p
Quantitative demands	1[a]	0.32	0.10	0.10	0.51	0.10	36.22	1	321	<0.001
	2[b]	0.36	0.13	0.13	0.50	0.03	10.51	1	320	0.001
Emotional demands	1[a]	0.16	0.03	0.02	0.53	0.03	8.76	1	320	0.003
	2[b]	0.19	0.03	0.03	0.53	0.01	2.50	1	319	0.120
Hiding demands	1[a]	0.25	0.06	0.06	0.52	0.06	21.90	1	321	<0.001
	2[b]	0.29	0.09	0.08	0.51	0.02	7.26	1	320	0.007
Sense of community	1[a]	0.34	0.11	0.11	0.50	0.11	40.28	1	319	<0.001
	2[b]	0.34	0.12	0.11	0.50	0.00	0.91	1	318	0.341
Quality of leadership	1[a]	0.29	0.09	0.08	0.51	0.09	29.79	1	321	<0.001
	2[b]	0.32	0.11	0.10	0.51	0.02	7.10	1	320	0.008

Table 4 Results of Hierarchical Regression Analyses for Psychological Strain: Variance Explained by the Generic Version Beyond the Contextualized Version.

Note. [a]Predictor: Contextualized version; [b]Predictors: Contextualized version and generic version. N = 323.

version contributed additional predictive value for the scales "quantitative demands," "hiding emotions," and "quality of leadership." Taken together, these findings suggest that while the contextualized scales offered incremental validity for certain constructs, the generic scales also accounted for unique variance for other scales, providing only partial support for H1b.

In the last step, we tested whether contextualization affected participants' reactions. Overall, participants' reactions were rather neutral (see Table 5). Table 5 also shows the t-values for paired tests as well as effect sizes: Liking and face validity did not significantly differ; however, perceived predictive validity differed between the generic and the contextualized versions of the stressor and resource scales, $t(312) = -2.39$, $p < 0.05$, $d = 0.14$. Thus, H2 was partly supported.

REACTION VARIABLE	$M\ (SD)_{generic}$	$M\ (SD)_{contextualized}$	t	p	d_{within}
Liking	3.14 (0.90)	3.21 (0.92)	−1.44	0.15	0.08
Face validity	3.54 (0.83)	3.54 (0.80)	−0.14	0.89	0.01
Perceived predictive validity	3.09 (0.81)	3.20 (0.81)	−2.39	0.02*	0.14

Table 5 Descriptive Statistics and Paired t-tests for Participant Reactions.

Note. N = 323.

*$p < 0.05$.

DISCUSSION

Our results indicate that complete contextualization can positively influence criterion validity even in work characteristics survey items, as the contextualized versions of three out of five scales explained additional variance in the criterion (i.e., psychological strain) beyond the generic versions. At the same time, contextualization did not lead to improved reliability or more favorable participant reactions and generic scales also seemed to account for unique variance in some scales, suggesting that the contextualizing work characteristics items may be more complex than initially expected and might not yield the results researchers had hoped for.

As contextualization effects have been primarily studied in personality questionnaires (e.g., Lievens, De Corte, & Schollaert 2008), this research aimed to investigate whether the positive effects of complete contextualization—rewriting items entirely to fit a specific context—that have been reported for personality scales (Holtrop et al. 2014) can also be observed in assessments of work characteristics. This study offers (only) weak support for this idea, as complete contextualization increased criterion-related variance in three out of five scales, suggesting that contextualization might help reduce the interpretative ambiguity often encountered by respondents to work characteristics scales (Greulich et al. 2021). In addition, participants reported higher perceived predictive validity for work performance with the contextualized items, indicating that the medical assistants (i.e., our participants) believe that valid inferences about their job performance can be made based on their perceptions of stressors and resources.

Nevertheless, the data only offered partial support, and where we found support, it was not as strong as hoped. Several caveats must be acknowledged. First, compared to studies on personality (e.g., Bing et al. 2004; Holtrop et al. 2014; Reddock, Biderman, & Nguyen 2011), the increase in criterion validity observed in this study was relatively modest, aligning more with studies that did not find validity increases (e.g., Holtrop, Born, & de Vries

2014). Second, identifying suitable items for complete contextualization specific to our context (i.e., medical assistants) was more challenging than anticipated, which limited us to a small subset of items. Consequently, the number of items per scale was fairly low, which led to reduced internal consistency for some scales. This reliability issue might have also obscured potential differences between the generic and contextualized versions. Third, the completely contextualized items were considerably longer than their generic counterparts, and the increased length might have made the items less appealing, and the latter effect might have masked an increased liking for the contextualized versions of the questionnaire. Fourth, the generic scales also accounted for unique variance in some cases, implying that both scale versions might explain different parts of the criterion variance. Fifth, the standard level of work-focused wording (such as in COPSOQ) may already be sufficient for capturing work characteristics. Sixth, we used a generic measure of psychological strain (i.e., the GHQ-12), which does not directly reference work-related content. It is therefore possible that the ambiguous findings regarding incremental validity reflect, at least in part, a mismatch in the level of specificity between predictors (which were work-related) and the outcome. Using more work-specific strain measures in future research may clarify these associations. Seventh, from a practical perspective, one could argue that creating fully contextualized items is such a tedious task with little guarantee of significant improvement that it may not be worth the effort.

In conclusion, our results tentatively suggest that completely contextualizing work characteristics scales may enhance the criterion validity of these measures. However, complete contextualization did not improve reliability or participants' reactions, and it also increased item length, raising doubts about its overall effectiveness. Therefore, further research on the contextualization of work characteristics surveys seems warranted.

DATA ACCESSIBILITY STATEMENT

Data available at https://osf.io/8k6vw/?view_only=ca5ce19ae2ab4c8da13d0fa6a30f7689.

FUNDING INFORMATION

German Research Foundation (to Cornelius J. König), Grant KO4244/4-1.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

BG played a lead role in conceptualization and supervision and an equal role in investigation, methodology, project administration, and writing. CK played an equal role in conceptualization, investigation, methodology, supervision, and writing. CM played a lead role in project administration and an equal role in conceptualization, investigation, and methodology.

AUTHOR AFFILIATIONS

Berit Greulich orcid.org/0000-0002-5955-7423
Industrial and Organizational Psychology, Saarland University, Saarland, Germany

Cornelius J. König orcid.org/0000-0003-0477-8293
Industrial and Organizational Psychology, Saarland University, Saarland, Germany

Carla V. Meixensberger
Industrial and Organizational Psychology, Saarland University, Saarland, Germany

REFERENCES

Ajzen, I., & **Fishbein, M.** (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*(5), 888–918. https://doi.org/10.1037/0033-2909.84.5.888

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

Credé, M., Bashshur, M., & **Niehorster, S.** (2010). Reference group effects in the measurement of personality and attitudes. *Journal of Personality Assessment, 92*(5), 390–399. https://doi.org/10.1080/00223891.2010.497393

Debus, M. E., Greulich, B., König, C. J., & **Kleinmann, M.** (2019). Insecure about how to rate your job insecurity? A two-study investigation into time frames applied to job insecurity measures. *Occupational Health Science, 3*(4), 421–435. https://doi.org/10.1007/s41542-019-00049-x

Eatough, E. M., & **Spector, P. E.** (2013). Quantitative self-report methods in occupational health psychology research. In M. Wang, R. Sinclair, & L. E. Tetrick (Eds.), *Research methods in occupational health psychology: Measurement, design, and data analysis* (pp. 248–267). Routledge.

Feldt, L. S., Woodruff, D. J., & **Salih, F. A.** (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*(1), 93–103. https://doi.org/10.1177/014662168701100107

Fryers, T., Brugha, T., Morgan, Z., Smith, J., Hill, T., Carta, M., Lehtinen, V., & **Kovess, V.** (2004). Prevalence of psychiatric disorder in Europe: The potential and reality of meta-analysis. *Social Psychiatry and Psychiatric Epidemiology, 39*(11), 899–905. https://doi.org/10.1007/s00127-004-0875-9

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*(4), 694–734. https://doi.org/10.2307/258595

Goldberg, D. P., & **Hillier, V. F.** (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine, 9*(1), 139–145. https://doi.org/10.1017/s0033291700021644

Greulich, B., Debus, M. E., Kleinmann, M., & **König, C. J.** (2021). Response behavior in work stress surveys: A qualitative study on motivational and cognitive processes in self- and other-reports. *European Journal of Work and Organizational Psychology, 30*(1), 40–57. https://doi.org/10.1080/1359432X.2020.1812580

Greulich, B., König, C. J., & **Fischer, N.** (2024). How relative is stress? The influence of social comparisons when responding to work stress surveys. *Journal of Personnel Psychology, 23*(1), 1–11. https://doi.org/10.1027/1866-5888/a000330

Havermans, B. M., Brouwers, E. P. M., Hoek, R. J. A., Anema, J. R., van der Beek, A. J., & **Boot, C. R. L.** (2018). Work stress prevention needs of employees and supervisors. *BMC Public Health, 18*(1), 642. https://doi.org/10.1186/s12889-018-5535-1

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

Holtrop, D., Born, M. P., & **de Vries, R. E.** (2014). Predicting performance with contextualized inventories, no frame-of-reference effect? *International Journal of Selection and Assessment, 22*(2), 219–223. https://doi.org/10.1111/ijsa.12071

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 and Assessment, 13*(1), 75–86. https://doi.org/10.1111/j.0965-075X.2005.00301.x

Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly, 24*(2), 285–308. https://doi.org/10.2307/2392498

Kristensen, T. S., Hannerz, H., Høgh, A., & **Borg, V.** (2005). The Copenhagen Psychosocial Questionnaire—a tool for the assessment and improvement of the psychosocial work environment. *Scandinavian Journal of Work, Environment & Health, 31*(6), 438–449. https://doi.org/10.5271/sjweh.948

Lautenschlager, G. J., & **Meade, A. W.** (2008). AlphaTest: A Windows program for tests of hypotheses about coefficient alpha. *Applied Psychological Measurement, 32*(6), 502–503. https://doi.org/10.1177/0146621607312307

Lazarus, R. S., & **Folkman, S.** (1984). *Stress, appraisal, and coping*. Springer.

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

Linden, M., Maier, W., Achberger, M., Herr, R., Helmchen, H., & **Benkert, O.** (1996). Psychische Erkrankungen und ihre Behandlung in Allgemeinarztpraxen in Deutschland: Ergebnisse aus einer Studie der Weltgesundheitsorganisation (WHO) [Psychological disorders and their treatment in general practice in Germany: Results of a WHO study.]. *Der Nervenarzt, 67*(3), 205–215.

Nübling, M., Stößel, U., Hasselhorn, H. M., Michaelis, M., & **Hofmann, F.** (2005). *Methoden zur Erfassung psychischer Belastungen: Erprobung eines Messinstrumentes (COPSOQ) [Methods for recording mental stress: Testing of a measuring instrument (COPSOQ)]*. Wirtschaftsverlag NW.

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

Sackett, P. R., Zhang, C., Berry, C. M., & **Lievens, F.** (2022). 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

Schlotzhauer, A. E., Ng, M. A., & **Su, S.** (2025). How to frame the frame of reference: A comparison of contextualization methods. *Journal of Business and Psychology, 40*(2), 385–403. https://doi.org/10.1007/s10869-024-09953-8

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

Schulze, 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

Semmer, N. K., Zapf, D., & **Dunckel, H.** (1995). Assessing stress at work: A framework and an instrument. In O. Svane & C. Johansen (Eds.), *Work and health: Scientific basis of progress in the working environment* (pp. 105–113). Office for Official Publications of the European Communities.

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

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & **Stoffey, R. W.** (1993). Applicant reactions to selection procedures. *Personnel Psychology, 46*(1), 49–76. https://doi.org/10.1111/j.1744-6570.1993.tb00867.x

Sonnentag, S., Tay, L., & **Nesher Shoshan, H.** (2023). A review on health and well-being at work: More than stressors and strains. *Personnel Psychology, 76*(2), 473–510. https://doi.org/10.1111/peps.12572

Tourangeau, R. (1987). Attitude measurement: A cognitive perspective. In H.-J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 149–162). Springer. https://doi.org/10.1007/978-1-4612-4798-2_8

Wiechmann, D., & **Ryan, A. M.** (2003). Reactions to computerized testing in selection contexts. *International Journal of Selection and Assessment, 11*(2–3), 215–229. https://doi.org/10.1111/1468-2389.00245

Wojujutari, A. K., Idemudia, E. S., & **Ugwu, L. E.** (2024). The evaluation of the General Health Questionnaire (GHQ-12) reliability generalization: A meta-analysis. *PLoS ONE 19*(7), e0304182. https://doi.org/10.1371/journal.pone.0304182

TO CITE THIS ARTICLE:
Greulich, B., König, C. J., & Meixensberger, C. V. (2025). Can We Improve the Assessment of Work Characteristics by Completely Contextualizing Items? *Scandinavian Journal of Work and Organizational Psychology*, 10(1): 13, 1–10. DOI: https://doi.org/10.16993/sjwop.348

Submitted: 18 September 2024 **Accepted:** 01 July 2025 **Published:** 29 August 2025