

How much is stress relative?



The influence of social comparisons when responding to work stress surveys

Berit Greulich, Cornelius J. König, and Nancy Fischer

Universität des Saarlandes, Germany

This article may not exactly replicate the final version published in *Journal of Personnel Psychology*. It is not the version of record and is therefore not suitable for citation.

Author Note

Berit Greulich  <https://orcid.org/0000-0002-5955-7423>, Cornelius J. König  <https://orcid.org/0000-0003-0477-8293>, and Nancy Fischer, Industrial and Organizational Psychology, Saarland University, Saarland, Germany.

Data available on request due to privacy/ethical restrictions. We have no conflicts of interest to disclose. We gratefully acknowledge funding from the German Research Foundation (to Cornelius J. König), Grant KO4244/4-1.

Correspondence concerning this article should be addressed to Berit Greulich, Saarland University, Campus A1 3, 66123 Saarbrücken, Germany. E-mail: berit.greulich@gmail.com

Abstract

In self-reports, employees frequently use self-selected social comparisons to assess workplace stressors and resources, but these comparisons vary within and between individuals. This study investigates how standardizing social comparison processes by adding a prescribed comparison to each item affects the reliability and validity of self-report scales measuring work-related stressors and resources, and how the standardized comparison affects scale means. A total of 208 employees were randomly assigned to one of two groups, comparing their perceptions to either their direct colleagues or without instructed comparison. Results indicate no effect on reliability, improvement in validity for one scale, and differences in means between groups for stressors and a resource scale. These findings suggest potential benefits and drawbacks of standardized social comparisons in self-report measures.

Keywords: work stress survey, social comparison, contextualization, reliability, validity

How much is stress relative?

The influence of social comparisons when responding to work stress surveys

Work stress surveys typically contain stressors (e.g., time pressure) and resources (e.g., task control) that should be rated regarding their frequency or intensity (e.g., Kristensen et al., 2005; Morgeson & Humphrey, 2006). Employees typically do this themselves in the form of self-report questionnaires, which is convenient and economical because a large number of employees can participate and are thus able to provide important information about their workplaces (Spector & Eatough, 2013). As noted by Greulich and colleagues (2021), employees use various social comparisons to evaluate their work-related stress, which might be a source of response bias. Given the diversity of social comparison objects, it might be an alternative to standardize the process of social comparison in order to investigate its impact on the reliability and validity of work stress surveys. Thus, the current study aims to investigate the potential effects of standardizing social comparison on the assessment of stressors and resources in the workplace with a focus on the psychometric implications of these findings. Additionally, the study will examine the role of social comparison in relation to work-related stress and how this impacts employees' evaluations.

Theoretical Background

In the context of occupational health, job characteristics have been a particular focus of stress theories. The Job Demand Control Model was originally developed to explain the impact and interaction of job demands and job control as a resource on employee health and well-being (Karasek, 1979; Karasek & Theorell, 1990). Since social support from colleagues and supervisors also play an important role in promoting employee health and well-being, the Job Demand Control Model has been expanded to include social support as a key resource (Johnson

& Hall, 1988; Van der Doef & Maes, 1999). However, to fully capture the complex and multifaceted nature of work-related stress, it may still be too limited (de Jonge et al., 2010). In response to this limitation, researchers have proposed the Job Demand Resource (JDR) model, which offers a more comprehensive and nuanced understanding of work-related stress (Demerouti et al., 2001). This model posits that all job characteristics can be categorized as either demands or resources. Demands (in the survey instrument used here, demands correspond to stressors, which is why this term is used in the following, Irmer et al., 2019) refer to the physical, psychological, social, and organizational aspects of work that require sustained effort and are associated with physical and psychological costs, such as burnout, fatigue, and job strain. Resources refer to aspects of work that are functional in achieving work goals, reduce job stressors, and are associated with positive outcomes, such as work engagement, job satisfaction, and well-being. As long as employees have sufficient resources, they can compensate for some stressors; however, if this balance is disrupted, stressors often lead to dissatisfaction, turnover intentions, and negative safety outcomes in the short term, as well as health detriments and sickness absence in the long term (e.g., Bakker et al., 2023). By focusing on job characteristics as stressors and resources, organizations can develop strategies to reduce stress and promote employee well-being. This might include adjusting workloads, providing opportunities for skill-building and career advancement, and creating a supportive work environment that values social connections and work-life balance. It is also important for organizations to consider a broader range of job characteristics when developing strategies to promote occupational health and well-being and to create a positive and inclusive work culture that values employee well-being. In particular, if organizations are interested in taking steps to reduce stressors and strengthen resources, a good starting point will be the assessment of work-related stressors and resources.

A large number of questionnaires have already been developed to assess stressors and resources at work (e.g., Kristensen et al., 2005; Morgeson & Humphrey, 2006). In many cases, these are self-reports of the frequency or intensity of stressors and resources evaluated on a five- to seven-point Likert scale. Self-reports are popular in this context because they are convenient and economical for organizations. Furthermore, employees likely know their workplace better than any other potential rater (Semmer et al., 1995). Typical stressors measured with such questionnaires are time pressure, concentration demands, and one-sidedness of physical stressors; typical resources that are surveyed are task control, support from colleagues or supervisors, and development opportunities. Many of these questionnaire items are phrased in general terms and can therefore be applied to many different workplaces. In addition to the causes of psychological strain, some questionnaires also include the potential consequences of stressors such as exhaustion, turnover intentions, and psychosomatic complaints (e.g., Kristensen et al., 2005).

Self-report questionnaires are widely used in psychology, but debates about measurement issues are prevalent in the literature (e.g., Einola & Alvesson, 2021). One challenge identified within this debate is the failure to consider the continuum that defines theoretical constructs, which can lead to incomplete measurement and compromised validity evidence. For example, Tay and Jebb (2018) propose a continuum specification approach to create an appropriate measure. Another challenge is to determine which kind of scaling is appropriate. For example, the General Health Questionnaire's (GHQ) traditional binary scoring system fails to distinguish between individuals with chronic symptoms and those without psychological distress (Whaley et al., 2005). Scaling issues matter even more if researchers create ratio variables as, for instance, in the case of effort-reward imbalance (Lang et al., 2020). Furthermore, there is an ongoing debate

about whether stress researchers should control for personality. This debate tends to center around the trait of negative affectivity (e.g., Debus et al., 2015): Whereas some researchers recommend controlling for NA bias, Spector and colleagues (2000) argue that NA should not be routinely partialled out, as it may play a substantial role in the job stress process. In addition, stress surveys might be affected by response styles such as acquiescence (e.g., Lee et al., 2020). This study aims to contribute to these debates on measurement issues in the well-being literature by providing a better understanding of work stress measurement with particular reference to social comparison processes.

Social comparisons in self-reports

Accurately completing self-reports can be a challenge, and it requires several cognitive processes for the respondents (Greulich et al., 2021; Tourangeau, 1987). Respondents have to understand the question asked of them, to activate mental representations of their job, and to compare these representations to the scale anchors. Therefore, they can experience some uncertainty about whether “sometimes” is the right answer to a question like “Can you decide for yourself how you want to do your job?”.

In such a situation of uncertainty, Festinger's (1954) theory of social comparisons predicts that answers are based on a comparison with others. Generally speaking, uncertainty about aspects of the self (e.g., abilities, traits, and feelings) increases the probability that social comparison processes will be used to make an evaluation (Festinger, 1954; Gerber et al., 2018). In most cases, the use of social comparison processes for assessment happens unintentionally and is not a conscious decision. Therefore, self-perceptions can be influenced by unaware cognitive processes like social comparisons (Heine et al., 2002). The use of social comparison processes by employees when filling out work stress surveys may be unproblematic when measuring

subjective stress levels. However, a potential problem could arise when many very different comparison objects are used to assess one's stress level. As Greulich et al. (2021) found, the comparison objects vary significantly both within a person and between individuals. As a result, interindividual and intraindividual differences may exist when using individual social comparisons to answer survey items. The inconsistent use of social comparison processes when answering questionnaires could thus affect the measurement accuracy of questionnaires. Consequently, the addition of a specified comparison object to standardize the social comparison process may positively impact psychometric properties, such as the reliability and validity, of work stress surveys.

Research on personality questionnaires has already shown that standardization of a frame-of-reference has positive effects on measurement accuracy (e.g., Lievens et al., 2008), and researchers in cross-cultural psychology who have been interested in understanding answers to self-reports in different cultures (e.g., Heine et al., 2008; Peng et al., 1997; Song et al., 2019) found that respondents self-rating on a Likert scale is influenced by social comparison, which they refer to as the reference-group effect. For instance, when Chinese people rate their level of individualism on a Likert scale they relate their response to the level of individualism of a salient reference group of other Chinese people and not on their absolute level of individualism (Heine et al., 2002). There is also preliminary evidence that this reference-group effect can be found in the assessment of personality: Credé and colleagues (2010) added four versions of an explicit reference group (e.g., their immediate family and people of the same age and gender) on items of a personality inventory. These explicit comparisons substantially influenced personality scores.

Social comparisons in work-related stressor and resource scales

Festinger's (1954) social comparison theory can also be applied to self-reports of work-related stressors and resources. Although this argument has not been empirically tested, it has been mentioned by several authors. For example, Semmer and colleagues (1995) noted that the perception of stressors and resources at work may be influenced by “social comparisons like ‘I am much better (worse) off than many of my colleagues’” (p. 105), and Buunk and Gibbons (2007) stated that they expect people to compare themselves more often in situations of high stressors. In addition, the results of a recent qualitative study indicate that employees use social comparison processes when filling out work stress surveys. Participants reported that they used different types of social comparisons (e.g., comparisons with direct co-workers, with people of their organization but with different tasks, and with different professions), that partly vary depending on the item (Greulich et al., 2021). Furthermore, colleagues tend to compete against each other when there are few opportunities for promotion, status, and recognition (Baumann et al., 2019; Prendergast, 1999), and situations that foster competition are likely to promote interest in social comparison for many people (Garcia et al., 2013).

Effects of item standardization on reliability and validity of work-related stressor and resource scales

Individuals tend to employ diverse social comparisons when completing work stress questionnaires (Goodman & Haisley, 2007; Greulich et al., 2021). On the one hand, this phenomenon pertains to the individual respondent, whose choice of reference group may vary depending on the item being evaluated, thereby resulting in a greater degree of intraindividual variability. On the other hand, it applies to all respondents, as each individual engages in individualized comparisons, ultimately resulting in a greater degree of interindividual variability. To address this issue and standardize the interpretation of items, it has been proposed in the

literature on personality questionnaires that items should be standardized (Schmit et al., 1995). While items lacking a context are more open to interpretation (Robie et al., 2000), by standardizing an item through the inclusion of a specific context (e.g., "at school" added to every personality item), respondents are expected to make consistent contextual references for each item, thus reducing intraindividual variance. Consequently, this reduction in variance is assumed to increase the reliability of the inventory, a proposition that has received empirical support in prior personality studies (Lievens et al., 2008; Reddock et al., 2011; Swift & Peterson, 2019).

Applying this concept to work stress surveys and considering respondents' inclination towards individual social comparisons, it is anticipated that item standardization through the inclusion of a social comparison object will reduce intraindividual variance in scales assessing work-related stressors and resources resulting in higher reliability. To examine this hypothesis, we have selected direct colleagues as the comparison group, as employees are presumed to be familiar with their colleagues to a relatively high degree, thus making such comparisons relatively effortless for respondents. It can be formally stated as follows:

Hypothesis 1: The reliabilities of work-related stressor and resource scales with an explicit comparison to a direct colleague are higher than the reliabilities of work-related stressor and resource scales without an explicit comparison to a direct colleague.

According to both classical and modern test theory, an increase in reliability leads to a corresponding increase in validity (e.g., Crocker & Algina, 1986). As a result of standardizing the social comparison process, it can be expected that all respondents will similarly approach the items, leading to a reduction in interindividual variance. This, in turn, could be accompanied by an increase in validity. Empirical studies investigating contextualized personality items have supported this argument (e.g., Lievens et al., 2008; Swift & Peterson, 2019; Voss et al., 2023).

The current approach aims to improve criterion validity by standardizing the items in work stress surveys through the incorporation of a specific social comparison (in this case, a direct colleague). Although the use of the term "close colleague" may still lead to individual variation in the selection of comparison objects, it is anticipated that this specification will nevertheless result in standardized item interpretation, as respondents are instructed to compare themselves to an individual with the same job responsibilities in the same work environment. Consequently, while the specific colleague may vary across respondents, the comparison object meets the same criteria for all respondents.

However, there is also evidence suggesting that the use of reference groups in stress questionnaires may not necessarily improve criterion validity. Specifically, Credé et al. (2010) found that the more specific the reference group, the lower the criterion validity. This suggests that individuals may have different perceptions of the specified reference group, leading to distinct scaling standards and potential bias in results. The greater the degree of specificity with which a comparison object is defined, the greater the likelihood that idiosyncratic individual characteristics such as behavior patterns or salient personality traits will shape the social comparison process. Such distinctive attributes of the comparison object can give rise to unique perceptions of the object that may vary significantly across individuals. Therefore, comparisons with a direct colleague could result in responses that are too specific referring to the content of the construct, which might reduce validity. Given that there are arguments for both sides, we formulate a research question.

Research Question 1: What is the impact of adding an explicit social comparison in stressor and resource items on the validity of work stress surveys?

The direction of social comparisons in work-related stressor and resource scales

In addition to the underlying motivations that prompt individuals to engage in social comparisons, the direction of the comparison represents a significant distinction within the existing body of research on social comparisons. Comparisons can be classified as either upward (where the target of comparison is perceived as superior) or downward (where the target of comparison is perceived as inferior). Downward comparisons have been found to evoke more positive emotions by enhancing self-esteem, whereas upward comparisons tend to elicit more negative emotions (Brickman & Bulman, 1977; Buunk & Gibbons, 2007; Taylor et al., 1996; Wood, 1989). Depending on the specific context, an upward comparison can also serve as a source of motivation for individuals to improve their situation (Collins, 1996; Taylor & Lobel, 1989). While Festinger's (1954) theory does not explicitly address the direction of social comparisons (i.e., whether individuals engage in downward or upward comparisons), it is pertinent to consider this aspect within the context of work-related stress.

The Stress as Offense to Self theory (SOS, Semmer et al., 2007, 2019) offers a framework that can provide insights into upward and downward social comparisons in the context of work-related stress. According to the SOS theory, stress arises when individuals perceive a threat to their sense of self. This occurs when demands at work (or in other life domains) exceed the individual's ability to cope, and when the individual perceives these demands as a challenge to their personal values or competencies. The SOS theory posits that stress is not solely the result of external demands or objective conditions, but is also influenced by subjective perceptions and evaluations. When individuals experience stress, they are motivated to restore their sense of self by reducing the perceived threat to their self-image.

According to the SOS theory (Semmer et al., 2007, 2019), employees strongly identify with their professional roles, which become integral to their identity and self-image. The maintenance of a positive personal and social self-image is considered a fundamental psychological need, and job performance and success assessments are particularly relevant to personal self-esteem. High levels of stressors at work may indicate that the employee is overwhelmed with tasks and lacks resources or support, leading to feelings of diminished worth and threatening self-esteem. As such, experiencing high stressors and low resources could undermine professional identity, and one potential coping strategy is to present oneself as being less stressed and possessing greater resources in comparison to a colleague. Consequently, employees may engage in downward social comparisons when evaluating their stressors and resources, comparing themselves to individuals with worse work conditions to maintain a positive self-image. At the same time, employees experiencing high stressors and low resources may fear that their performance will decline in response to stress, which could be observed by others. To protect their self-esteem and justify any performance drops, they may resort to upward comparisons, rating their own stressors as much higher and their resources as much less pronounced than is the case with their colleagues, suggesting that their heightened stress is a result of having more responsibilities or tasks compared to their colleagues. Based on the argument presented by the SOS theory (Semmer et al., 2007, 2019), it is possible that employees use social comparisons to protect their self-esteem when facing high work stress. If the social comparison is predominantly downward directed, employees may report lower average stressors and higher average resources compared to a group without an explicit comparison. Conversely, if the social comparison is predominantly upward directed, employees may report higher average stressors and lower average resources. It remains unclear whether an explicit social comparison,

specifically with colleagues in our case, is more likely to trigger a downward or an upward comparison. Thus, the following research question is formulated:

Research Question 2: Does the incorporation of an explicit comparison in work stress survey items elicit a predominantly downward comparison, resulting in lower mean stressor values and higher mean resource values relative to the control group, or an upward comparison, leading to higher mean stressor values and lower mean resource values relative to the control group?

Method

Participants

We recruited employees from nine rehabilitation hospitals in Germany by sending out 600 paper-and-pencil questionnaires to employees. The final sample consisted of 208 employees (i.e., a response rate of 34.7%) and was mostly female (78.8%). The majority of the participants (49.5%) were over 50 years old, 24.5% were between 41 and 50 years old, 16.7% were 30 to 40 years old, and 9.3% were under 30 years old. The participants' professional fields primarily consisted of facility management (30.0%) and healthcare (21.0%). Other fields were psychotherapy (20.7%), public administration (13.5%), and childcare (14.8%). The control ($n = 105$) and experimental ($n = 103$) groups did not differ in terms of gender, $F(1, 183) = .006, p = .940$, age, $F(1, 183) = .740, p = .391$, language, $F(1, 183) = .06, p = .806$, professional fields, $F(1, 183) = 2.02, p = .157$, and job satisfaction, $F(1, 183) = .510, p = .476$. Participation was voluntary.

Procedure and experimental manipulation

In a between-subjects design, participants were randomly assigned to one of two groups: The control group ($n = 105$) was given the original questions without changes in the wording

(e.g., “How often are you under time pressure?”) and instructed to evaluate the stressors and resources in their workplace; The items of the experimental group ($n = 103$) were manipulated by adding the wording “in comparison to your close colleague” at the end of the same items (e.g., “How often are you under time pressure compared to your close colleague?”). Participants placed their questionnaires in a ballot box that was emptied after approximately three weeks, and the questionnaires were then sent back to us.

Measures

Work stress survey. Four scales were used to assess typical stressors: time pressure, work environment, one-sided physical stressors, and work-life balance. Two scales were used to assess typical resources at work: task control and development opportunities/meaning of work. The scales were taken from two common German work stress surveys: the “Instrument for Stress-Oriented Task Analysis” (ISTA; Irmer et al., 2019; Semmer et al., 1999) and the “Copenhagen Psychosocial Questionnaire” (COPSOQ; Kristensen et al., 2005; Nübling et al., 2005). The criteria for selection were that the stressors and resources should be relevant for all occupational groups at a hospital and that adding the comparison should be feasible without changing items. The items measure stressors and resources either in terms of frequency or intensity on a five-point Likert scale ranging from 1 (disagree strongly/never) to 5 (agree strongly/always; for details see Table 1). However, the scales ‘one-sided physical stressors’ and ‘work-life balance’ yielded low reliability (Cronbach’s $\alpha = .23$ and $.43$ respectively) and were therefore excluded from further analysis.

Job satisfaction. To test the effects of an explicit comparison with colleagues on validity, we used job satisfaction as a criterion, assessed with the Neuberger and Allerbeck (1978) job satisfaction scale, which is a widely used job satisfaction scale in Germany (see e.g., Maier &

Brunstein, 2001). Participants rate satisfaction with their job, working conditions, the relationship with their supervisor, the relationship with their colleagues, promotion opportunities, organization and management, and benefits and pay. The participants indicate their level of agreement with the seven items on a five-point Likert scale ranging from 1 (disagree strongly) to 5 (agree strongly).

Control variables. Gender and age have been shown to influence ratings of stressors and resources at work (e.g., Matud, 2004; Shultz et al., 2010; Watson et al., 2011), and we thus controlled for both variables.

Stress mindset. To additionally check whether participants perceive stress as something rather negative, their general attitude toward stress was assessed with the following four items from the Stress Mindset Measure (Crum et al., 2013, in its German translation, Schollmeyer, 2004): (1) “The effects of stress are negative and should be avoided.”, (2) “Experiencing stress increases my performance and productivity.”, (3) “Experiencing stress affects my health and vitality.”, (4) “The effects of stress are positive and should be used.” Respondents indicate their level of agreement with a five-point Likert scale ranging from 1 (disagree strongly) to 5 (agree strongly), and items (2) and (4) are reverse coded. Therefore, a high score indicates a negative attitude towards stress.

Results

Preliminary analyses

Participants had on average a fairly negative attitude towards stress, as indicated by high Stress Mindset Measurement scale values in both groups (the group with a colleague comparison: $M = 4.11$, $SD = 0.82$; the group without a colleague comparison: $M = 4.17$, $SD = 0.75$). The two means were not significantly different from each other, $t(206) = .52$, $p = .60$).

Test of the hypothesis and the research questions

To test for reliability differences, we calculated Cronbach's α for each scale and each condition and compared them using the computer program Alpha Test (Lautenschlager & Meade, 2008). Contrary to Hypothesis 1, Cronbach's α s did not significantly differ between groups (see Table 2).

Table 3 shows how well the stressors and resources with and without the comparison predicted job satisfaction. Correlation coefficients were compared for statistical significance (Lautenschlager & Meade, 2008). As shown in Table 5, only the 'work environment' stressor scale predicted job satisfaction better with social comparison vs. without comparison.

Because previous literature suggests that age and gender influence the assessment of stressors and resources, these variables were considered as covariates. The assumptions of a MANCOVA were met: An ANOVA showed the independence of the covariates and the treatment effect with $F(1, 202) = 0.96, p = .33$, for age and $F(1, 206) = 0.56, p = .46$, for gender. Furthermore, the assumption of homogeneity of regression slopes could be accepted for the stressor and resource scales. The means, adjusted means, standard deviations, and standard errors for the work-related stressor and resource scales are shown in Table 4.

The MANCOVA for the stressor scales (see Table 5) showed a significant overall difference between the groups with and without explicit social comparison, $F(2, 199) = 3.44, p < .05$. The tests of between-subjects effects were significant for both time pressure, $F(1, 200) = 4.60, p < .05$, and work environment, $F(1, 200) = 4.14, p < .05$. Therefore, the group with colleague comparisons had lower average stressor scores than the group without comparisons, supporting a downward comparison.

The MANCOVA for resources (see Table 5) showed a significant overall difference between the groups with and without explicit social comparison, $F(2, 199) = 11.67, p < .001$. The test of between-subjects-effects was significant for development opportunities, $F(1, 200) = 6.65, p < .05$, with a higher mean for the group with comparisons compared to the group without comparisons. The effect on task control did not reach conventional significance levels (and was descriptively even opposite to the hypothesized direction). Accordingly, these data do not allow a clear to be drawn.

Discussion

The purpose of this study was to investigate the impact of standardized social comparisons on the evaluation of work-related stressors and resources. Specifically, we tested the hypothesis that contextualizing items through the inclusion of a defined comparison object would affect the reliability of the measurement. Furthermore, as part of a research question, we were interested in investigating the impact of integrating a standardized social comparison process on the validity of assessing work-related stress. Although contextualization did not have a significant impact on reliability, it did improve the validity of one stressor scale by reducing interindividual error variance. Regarding the research question of the directionality of social comparison processes, the results indicated that for stressors, the group that included colleague comparisons had lower average stressor scores in comparison to the group without comparisons. Conversely, for resources, higher mean values were observed only for the development opportunities scale. Overall, our findings indicate that the standardization of social comparison in work-related stressors and resources had a discernible influence on the evaluation. However, the effects on the psychometric variables were not as substantial as anticipated. Nevertheless, this

approach presents a potential alternative to previous measurement techniques that requires further elucidation.

The standardized social comparison influenced the means of the scales *time pressure*, *work environment*, and *development opportunities*. This finding aligns with the results of Credé et al. (2010), who found that mean values of a scale for personality vary depending on the reference-group. Going beyond previous research in the personality realm, we derived from SOS theory (Semmer et al., 2007, 2019) that stress, if seen as something negative, threatens self-esteem and that employees tend to protect it. Our data show that employees have negative attitudes toward stress and tend to take downward comparisons to protect their self-esteem: Lower stressor values and higher resource values were observed in the group with peer comparison.

Based on studies in the personality domain (e.g., Reddock et al., 2011), we expected that the addition of a context in the form of a social comparison should increase scale reliability because the unification of the context should reduce *intraindividual* variance (Lievens et al., 2008). However, the data did not support this hypothesis, which implies a difference between personality test items and stressor/resources items: Whereas people might consider different comparison targets when filling out (e.g., a friend for extraversion item no. 1 vs. the sister for extraversion item no. 2), people filling out stressor and resources items do not seem to vary strongly enough in their comparison targets between items to make this measurable in the context of this study.

The incorporation of a targeted comparison was anticipated to improve the validity of the evaluation. However, it was found to only have a significant impact on one scale, namely the work environment scale. This finding may support the theoretical argument that standardizing

the context can reduce interindividual error variance (Lievens et al., 2008). In the absence of an explicit comparison, employees may compare themselves to other occupational groups, family members, or colleagues (see Greulich et al., 2021), which increases interindividual error variance.

Limitations and future research

As with any study, this work is subject to limitations. The choice of items from common work stress questionnaires was restricted by the manipulation (e.g., because items already had a different kind of comparison in their formulation, or their content did not allow us to add a comparison). Thus, the pool of potential scales available in the end was quite limited, and some scales consisted of only two to four items, which was likely the reason why their reliabilities were too low to use. Ideally, future research is able to work with other scales. Although the sample size in this study was rather small and maybe not be sufficient to detect true small differences in two independent correlations (Vanhove, 2015), we considered the investigation of the influence of social comparison processes on reliability and validity appropriate to get first insights into possible effects. A replication with a larger sample would also allow the use of other statistical techniques to investigate possible biases (e.g., confirmatory factor analyses or differential item analyses based on item response theory, Somaraju et al., 2022; Thissen et al., 1993, but see also Robitzsch & Lüdtke, 2022). Furthermore, all participants were hospital staff, and the majority of the participants were female. Although a female majority is typical for German hospital workers (see Brehm et al., 2021), future research should try to replicate these findings with other samples of employees. Care should then also be taken to ensure that the gender variable is more equally distributed than in our sample. Even though this study did not show any differences depending on gender, there are indications that men and women differ in

their use of downward comparisons to protect themselves (Kemmelmeier & Oyserman, 2001). In addition, the effects of adding a social comparison in this study were not as strong as initially expected, maybe because participants worked in hospitals where working conditions may not be as competitive as they would be in other sectors. Further research should also investigate whether there are other comparisons that are also relevant in the context of work-related stressors and resources. For example, Greulich et al. (2021) mentioned that employees might also compare themselves to other professions or to family members – or to themselves in a previous job.

Practical implications

This study implies that organizations should pay more attention to the details of administration when collecting survey data about workplace stressors and resources. For example, the text introducing a survey might trigger certain comparisons (or only certain comparisons for some employees but not others). Organizations could also attempt to standardize such comparison processes by adding an explicit comparison to the introductory text (or to each item). Furthermore, organizations should become more aware that data from standard stress surveys are also influenced by comparison processes. Thus, before taking such data too literally and before using it to initiate actions (e.g., workplace changes for a certain unit), it is likely beneficial to reach out to this unit and its employees and to try to understand the situation and the perception of the situation in this unit, maybe by using qualitative interviews.

Conclusion

In this study, we have contributed to the ongoing discourse surrounding measurement issues in the field of work stress research, which has a long-standing tradition of discussion and debate (e.g., Lang et al., 2020; Whaley et al., 2005). Through our examination of work stress measurement with a particular focus on social comparison processes, we have gained insights

into the complexities of measuring subjective experiences in the workplace. A notable finding of this study is that the inclusion of a comparison variable can influence the means of work stress survey scales, potentially reducing interindividual error variance. The reason why the reduction was not observed for all scales requires further investigation. Therefore, this study encourages researchers to continue to engage in these measurement debates and to develop innovative approaches to measurement that are better aligned with theoretical constructs, and thereby advance our understanding of the complex nature of human experiences in the workplace.

References

- Bakker, A. B., Demerouti, E., & Sanz-Vergel, A. (2023). Job Demands–Resources Theory: Ten years later. *Annual Review of Organizational Psychology and Organizational Behavior*, *10*(1), 25–53. <https://doi.org/10.1146/annurev-orgpsych-120920-053933>
- Baumann, O., Eggers, J. P., & Stieglitz, N. (2019). Colleagues and competitors: How internal social comparisons shape organizational search and adaptation. *Administrative Science Quarterly*, *64*(2), 275–309. <https://doi.org/10.1177/0001839218766310>
- Brehm, T. T., Schwinge, D., Lampalzer, S., Schlicker, V., Küchen, J., Thompson, M., Ullrich, F., Huber, S., Schmiedel, S., Addo, M. M., Lütgehetmann, M., Knobloch, J. K., Schulze zur Wiesch, J., & Lohse, A. W. (2021). Seroprevalence of SARS-CoV-2 antibodies among hospital workers in a German tertiary care center: A sequential follow-up study. *International Journal of Hygiene and Environmental Health*, *232*, 113671. <https://doi.org/10.1016/j.ijheh.2020.113671>
- Brickman, P., & Bulman, R. J. (1977). Pleasure and pain in social comparison. In L. M. Sulsky & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives*. (pp. 149–186). Hemisphere.
- Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes*, *102*(1), 3–21. <https://doi.org/10.1016/j.obhdp.2006.09.007>
- Collins, R. L. (1996). For better or worse: The impact of upward social comparison on self-evaluations. *Psychological Bulletin*, *119*(1), 51–69. <https://doi.org/10.1037/0033-2909.119.1.51>

- 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>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt Rinehart & Winston.
- de Jonge, J., van Vegchel, N., Shimazu, A., Schaufeli, W., & Dormann, C. (2010). A longitudinal test of the Demand–Control Model using specific job demands and specific job control. *International Journal of Behavioral Medicine, 17*(2), 125–133.
<https://doi.org/10.1007/s12529-010-9081-1>
- Debus, M. E., König, C. J., Kleinmann, M., & Werner, C. S. (2015). Examining the effects of negative affectivity on self- and supervisor ratings of job stressors: The role of stressor observability. *Work & Stress, 29*(4), 341–361.
<https://doi.org/10.1080/02678373.2015.1075233>
- Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology, 86*(3), 499–512.
<https://doi.org/10.1037/0021-9010.86.3.499>
- Einola, K., & Alvesson, M. (2021). Behind the numbers: Questioning questionnaires. *Journal of Management Inquiry, 30*(1), 102–114. <https://doi.org/10.1177/1056492620938139>
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140.
<https://doi.org/10.1177/001872675400700202>
- Garcia, S. M., Tor, A., & Schiff, T. M. (2013). The psychology of competition: A social comparison perspective. *Perspectives on Psychological Science, 8*(6), 634–650.
<https://doi.org/10.1177/1745691613504114>

- Gerber, J. P., Wheeler, L., & Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin*, *144*(2), 177–197. <https://doi.org/10.1037/bul0000127>
- Goodman, P. S., & Haisley, E. (2007). Social comparison processes in an organizational context: New directions. *Organizational Behavior and Human Decision Processes*, *102*(1), 109–125. <https://doi.org/10.1016/j.obhdp.2006.10.005>
- 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>
- Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science*, *19*(4), 309–313. <https://doi.org/10.1111/j.1467-9280.2008.02085.x>
- Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, *82*(6), 903–918. <https://doi.org/10.1037//0022-3514.82.6.903>
- Irmer, J. P., Kern, M., Schermelleh-Engel, K., Semmer, N. K., & Zapf, D. (2019). The Instrument for Stress-Oriented Task Analysis (ISTA). *Zeitschrift für Arbeits- und Organisationspsychologie*, *63*(4), 217–237. <https://doi.org/10.1026/0932-4089/a000312>
- Johnson, J. V., & Hall, E. M. (1988). Job strain, work place social support, and cardiovascular disease: A cross-sectional study of a random sample of the Swedish working population. *American Journal of Public Health*, *78*(10), 1336–1342. <https://doi.org/10.2105/ajph.78.10.1336>

- 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>
- Karasek, R. A., & Theorell, T. (1990). *Healthy work: Stress productivity and reconstruction of working life*. Basic Books.
- Kemmelmeier, M., & Oyserman, D. (2001). Gendered influence of downward social comparisons on current and possible selves. *Journal of Social Issues*, 57(1), 129–148.
<https://doi.org/10.1111/0022-4537.00205>
- Kristensen, T. S., Borritz, M., Villadsen, E., & Christensen, K. B. (2005). The Copenhagen Burnout Inventory: A new tool for the assessment of burnout. *Work & Stress*, 19(3), 192–207. <https://doi.org/10.1080/02678370500297720>
- Lang, J. W. B., Van Hoeck, S., & Runge, J. M. (2020). Methodological and conceptual issues in studying effort-reward fit. *Journal of Managerial Psychology*, 37(5), 498–512.
<https://doi.org/10.1108/JMP-11-2019-0659>
- 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>
- Lee, S., Vasquez, E., Ryan, L., & Smith, J. (2020). Measurement equivalence of subjective well-being scales under the presence of acquiescent response style for the racially and ethnically diverse older population in the United States. *Survey Research Methods*, 14(4), Article 4. <https://doi.org/10.18148/srm/2020.v14i4.7413>

- 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>
- Maier, G. W., & Brunstein, J. C. (2001). The role of personal work goals in newcomers' job satisfaction and organizational commitment: A longitudinal analysis. *Journal of Applied Psychology, 86*(5), 1034–1042. <https://doi.org/10.1037/0021-9010.86.5.1034>
- Matud, M. P. (2004). Gender differences in stress and coping styles. *Personality and Individual Differences, 37*(7), 1401–1415. <https://doi.org/10.1016/j.paid.2004.01.010>
- Morgeson, F. P., & Humphrey, S. E. (2006). The Work Design Questionnaire (WDQ): Developing and validating a comprehensive measure for assessing job design and the nature of work. *Journal of Applied Psychology, 91*(6), 1321–1339. <https://doi.org/10.1037/0021-9010.91.6.1321>
- Neuberger, O., & Allerbeck, M. (1978). *Messung und Analyse von Arbeitszufriedenheit: Erfahrungen mit dem "Arbeitsbeschreibungsbogen (ABB)" [Measurement and analysis of job satisfaction: experiences with the "job description questionnaire (ABB)"]*. Hans Huber.
- Nübling, M., Stöbel, U., Hasselhorn, H. M., Michaelis, M., & Hofmann, F. (2005). COPSOQ: Ein Instrument zur Messung psychischer Belastungen am Arbeitsplatz. *Gesundheitswesen, 67*, V60. <https://doi.org/10.1055/s-2005-920554>
- Peng, K., Nisbett, R. E., & Wong, N. Y. C. (1997). Validity problems comparing values across cultures and possible solutions. *Psychological Methods, 2*(4), 329–344. <https://doi.org/10.1037/1082-989X.2.4.329>

- Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 37(1), 7–63. <https://doi.org/10.1257/jel.37.1.7>
- 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., 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>
- Robitzsch, A., & Lüdtke, O. (2022). *Why measurement invariance is not necessary for valid group comparisons* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/cjyqp>
- 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>
- Schollmeyer, M. (2004). *Erstellung und Validierung der deutschen Version des Stress Mental Measure Testinventars [Development and validation of the German version of the Stress Mental Measure Inventory]* [Bachelorarbeit]. Saarland University.
- Semmer, N. K., Jacobshagen, N., Meier, L. L., & Elfering, A. (2007). Occupational stress research: The “stress-as-offence-to-self” perspective. In J. Houdmont & S. McIntyre (Eds.), *Occupational health psychology: European perspectives on research, and practices* (Vol. 2, pp. 42–66). ISMAI.

- Semmer, N. K., Tschan, F., Jacobshagen, N., Beehr, T. A., Elfering, A., Kälin, W., & Meier, L. L. (2019). Stress as Offense to Self: A promising approach comes of age. *Occupational Health Science*, 3(3), 205–238. <https://doi.org/10.1007/s41542-019-00041-5>
- 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.
- Semmer, N. K., Zapf, D., & Dunckel, H. (1999). Instrument zur stressbezogenen Tätigkeitsanalyse ISTA [Instrument for stress-oriented task analysis ISTA]. In H. Dunckel (Ed.), *Handbuch psychologischer Arbeitsanalyseverfahren* (pp. 179–204). vdf Hochschulverlag.
- Shultz, K. S., Wang, M., Crimmins, E. M., & Fisher, G. G. (2010). Age differences in the demand-control model of work stress: An examination of data from 15 European countries. *Journal of Applied Gerontology*, 29(1), 21–47. <https://doi.org/10.1177/0733464809334286>
- Somaraju, A. V., Nye, C. D., & Olenick, J. (2022). A review of measurement equivalence in organizational research: What’s old, what’s new, what’s next? *Organizational Research Methods*, 25(4), 741–785. <https://doi.org/10.1177/109442812111056524>
- Song, H., Cramer, E. M., & Park, N. (2019). Cultural differences in social comparison on Facebook. *Behaviour & Information Technology*, 38(2), 172–183. <https://doi.org/10.1080/0144929X.2018.1519037>

- Spector, P. E., & Eatough, E. (2013). Quantitative self-report methods in occupational health psychology research. In M. Wang, S. Sinclair, & L. E. Tetrick (Eds.), *Research methods in occupational health psychology* (pp. 248–267). Taylor & Francis.
- Spector, P. E., Zapf, D., Chen, P. Y., & Frese, M. (2000). Why negative affectivity should not be controlled in job stress research: Don't throw out the baby with the bath water. *Journal of Organizational Behavior*, *21*, 79–95. [https://doi.org/10.1002/\(SICI\)1099-1379\(200002\)21:1<79::AID-JOB964>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1099-1379(200002)21:1<79::AID-JOB964>3.0.CO;2-G)
- 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>
- Tay, L., & Jebb, A. T. (2018). Establishing construct continua in construct validation: The process of continuum specification. *Advances in Methods and Practices in Psychological Science*, *1*(3), 375–388. <https://doi.org/10.1177/2515245918775707>
- Taylor, S. E., & Lobel, M. (1989). Social comparison activity under threat: Downward evaluation and upward contacts. *Psychological Review*, *96*(4), 569–575. <https://doi.org/10.1037/0033-295X.96.4.569>
- Taylor, S. E., Wayment, H. A., & Carrillo, M. (1996). Social comparison, self-regulation, and motivation. In R. M. S. E. T. Higgins (Ed.), *Handbook of motivation and cognition, Vol. 3: The interpersonal context* (pp. 3–27). Guilford Press.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In *Differential item functioning* (pp. 67–113). Erlbaum. <https://doi.org/10.1075/z.62.13kok>

- 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.
- Van der Doef, M., & Maes, S. (1999). The Job Demand-Control (-Support) Model and psychological well-being: A review of 20 years of empirical research. *Work & Stress, 13*(2), 87–114. <https://doi.org/10.1080/026783799296084>
- Vanhove, J. (2015, April 14). *Power simulations for comparing independent correlations*. <https://janhove.github.io/design/2015/04/14/power-simulations-for-comparing-independent-correlations>
- Voss, N. M., Chlevin-Thiele, C., Lake, C. J., & Warren, C.-L. (2023). Using item response theory to understand the effects of scale contextualization: An illustration using decision making style scales. *International Journal of Testing, 23*(1), 34–55. <https://doi.org/10.1080/15305058.2022.2047692>
- Watson, S. B., Goh, Y. W., & Sawang, S. (2011). Gender influences on the work-related stress-coping process. *Journal of Individual Differences, 32*(1), 39–46. <https://doi.org/10.1027/1614-0001/a000033>
- Whaley, C. J., Morrison, D. L., Payne, R. L., Fritschi, L., & Wall, T. D. (2005). Chronicity of psychological strain in occupational settings and the accuracy of the General Health Questionnaire. *Journal of Occupational Health Psychology, 10*(4), 310–319. <https://doi.org/10.1037/1076-8998.10.4.310>
- Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological Bulletin, 106*(2), 231–248. <https://doi.org/10.1037/0033-2909.106.2.231>

Table 1

Work-Related Stressor and Resource Items, Chosen from the Instrument of Stress-oriented Task Analysis (ISTA, Semmer et al., 1999) and Copenhagen Psychosocial Questionnaire (COPSOQ, Nübling et al., 2005)

Scale	Item	Inventory (Number)
Stressors		
Time Pressure	How often are you pressed for time?	ISTA (ZD1)
	How often must you finish work later because you have too much to do?	ISTA (ZD4)
Work Environment	How often is a fast pace of work required of you?	ISTA (ZD6)
	How often do you have to do physically strenuous work?	COPSOQ (B8b-1)
	How often are you exposed to noise or loud background noise at your workplace?	COPSOQ (B8b-2)
	How often do you come in contact with chemicals or hazardous substances at your work?	COPSOQ (B8b-3)
One-sided Physical Stressors	How often are you exposed to extreme temperatures or a draft at your workplace?	COPSOQ (B8b-4)
	How often does your work activity require you to bend from the waist?	ISTA (EBA2)
	How often does your work activity require you to take a twisted or unusual posture?	ISTA (EBA3)
	How typical is sitting for long periods of time in your work activity?	ISTA (EBA4)
Work-life Balance	How typical is standing for long periods of time in your work activity?	ISTA (EBA5)
	I take care of work-related tasks outside of my working time as well.	COPSOQ (B2-5)
	I'm available in my free time for people I deal with professionally.	COPSOQ (B2-6)
Resources		
Task Control	Considering your workplace in general, how much can you change the sequence of your different tasks yourself?	ISTA (HS1)
	How much influence do you have on the work that is assigned to you?	ISTA (HS2)
	Considering your work activity in general, how much opportunity is there for you to make your own decisions?	ISTA (HS3)
	Can you yourself decide on which way to carry out your work?	ISTA (HS4)
Development Opportunities	Do you have the possibility of learning new things through your work?	COPSOQ (B5-1)
	Can you use your skills or expertise in your work?	COPSOQ (B5-2)
	Is your work meaningful?	COPSOQ (B5-3)

Table 2*Reliabilities for Each Group*

Scale	Cronbach's α		χ^2	p
	Without comparison ($n = 105$)	With comparison ($n = 103$)		
Stressors				
Time Pressure	.80	.74	.90	.34
Work Environment	.82	.79	.27	.61
Resources				
Task Control	.88	.87	.20	.66
Development Opportunities	.77	.81	.44	.51

Note. The χ^2 and p values describe the significance of the reliability differences.

Table 3*Criterion Validities for Each Group*

Scale	Correlations with job satisfaction		
	Without comparison	With comparison	<i>p</i>
	(<i>n</i> = 105)	(<i>n</i> = 103)	
Stressors			
Time Pressure	-.38	-.30	.24
Work Environment	-.25	-.47	.03*
Resources			
Task Control	.30	.37	.32
Development Opportunities	.42	.36	.32

Note. The *p* values describe the significance of the differences between correlations.

**p* < .05.

Table 4

Means, Adjusted Means, Standard Deviations, and Standard Errors of the Work-related Stressor and Resource Scales for the two Groups

		Without Comparison (<i>n</i> = 105)	With Comparison (<i>n</i> = 103)
Stressors			
Time Pressure	<i>M</i>	3.29	3.05
	<i>(SD)</i>	(0.80)	(0.82)
	<i>M_{adj}</i>	3.29	3.05
	<i>(SE)</i>	(0.08)	(0.08)
Work Environment	<i>M</i>	2.56	2.25
	<i>(SD)</i>	(1.17)	(0.94)
	<i>M_{adj}</i>	2.56	2.25
	<i>(SE)</i>	(0.10)	(0.11)
Resources			
Task Control	<i>M</i>	2.50	2.26
	<i>(SD)</i>	(1.00)	(0.87)
	<i>M_{adj}</i>	2.50	2.27
	<i>(SE)</i>	(0.09)	(0.09)
Development Opportunities	<i>M</i>	2.43	2.78
	<i>(SD)</i>	(.99)	(1.01)
	<i>M_{adj}</i>	2.43	2.79
	<i>(SE)</i>	(0.10)	(0.10)

Note. Adjustments for age and gender as control variables.

Table 5*Results of the MANCOVAs of the Stressor and Resource Scales and Their Respective Covariates*

	<i>F</i>	<i>p</i>	Partial η^2
Stressors			
Time Pressure	4.60	< .05*	.02
Work Environment	4.14	< .05*	.02
Age	0.73	.48	.01
Gender	0.41	.67	.00
Resources			
Task Control	3.08	.08	.02
Development Opportunities	6.65	< .05*	.03
Age	0.11	.90	.00
Gender	2.46	.09	.02

Note. * $p < .05$.