

RESEARCH

Open Access



Assessing the Turkish version of the Individual Work Performance Questionnaire (IWPQ) for its validity and reliability

Esma Özkan^{1*}, Sümeyye Belhan Çelik² and Linda Koopmans³

Abstract

Backgrounds This study aimed to assess the validity and reliability of the Turkish version of the Individual Work Performance Questionnaire (IWPQ).

Methods A total of 340 adult participants who had been working for at least 8 h a day completed the study questionnaire online. The questionnaire consisted of sections on demographics and socioeconomic characteristics, the IWPQ, and the ultrashortened version of the Utrecht Work Engagement Scale (UWES-3). Principal component analysis (PCA) with Promax rotation was conducted to assess the e three-factor structure for the Turkish version of the IWPQ. Cronbach's alpha was used to determine internal consistency. The intraclass correlation coefficient (ICC) was used to assess test-retest reliability. Spearman correlation coefficients were calculated between the IWPQ and UWES-3 scales to evaluate convergent validity.

Results The three factors accounted for a total variance ratio of 60.7%. Cronbach's alphas of three subscales (task performance, contextual performance, and counterproductive work behavior) were 0.894, 0.875, and 0.796, respectively. Test-retest reliability showed high agreement, with ICCs of 0.96, 0.98, and 0.91 for the respective scales. There was a statistically significant correlation between UWES-3 and the task performance subscale ($\rho = 0.367$ and $p < 0.001$), a statistically significant correlation between UWES-3 and the contextual performance subscale ($\rho = 0.403$ and $p < 0.001$), and a statistically significant correlation between UWES-3 and the CWB subscale ($\rho = -0.352$ and $p < 0.001$).

Conclusions The Individual Work Performance Questionnaire (IWPQ) identifies work performance. The Turkish version of the IWPQ scales demonstrated validity and reliability in assessing individual work performance.

Keywords Individual work performance scale, Turkish, Validity, Reliability

*Correspondence:

Esma Özkan
esmaa.ozkannn@gmail.com

¹Department of Occupational Therapy, Gülhane Faculty of Health Sciences, University of Health Sciences, Ankara, Türkiye

²Department of Occupational Therapy, Hamidiye Faculty of Health Sciences, University of Health Sciences, İstanbul, Türkiye

³Department of Sustainable Productivity and Employability, TNO Healthy Living and Work, Leiden, Netherlands



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Work performance is a critical aspect of organizational success, encompassing employees' behaviors, actions, and outcomes aligned with organizational goals [1]. The measurement of individual work performance is a critical component of organizational success, as it directly influences productivity, employee engagement, and overall workplace efficiency [2]. Koopmans et al. [1] described three dimensions of work performance: task performance, contextual performance, and counterproductive work behavior. Task performance refers to employees' competency and expertise in executing their jobs efficiently [3]. Contextual performance, on the other hand, pertains to employees' voluntary engagement in extra tasks, support for colleagues, and cooperation [4]. Counterproductive work behavior involves negative actions that hinder organizational performance and impede the achievement of objectives [4].

Some studies have confirmed a significant correlation between work performance and different health outcomes, highlighting the reciprocal connection between an individual's physical and mental well-being and their effectiveness in the workplace [5–8]. Studies suggest that employees with elevated stress levels and compromised mental well-being are more prone to reduced work performance, as stress detrimentally affects cognitive abilities such as attention, memory, and decision-making [9, 10]. Moreover, chronic health conditions such as cardiovascular disease and musculoskeletal disorders have been associated with decreased work capacity and productivity [11–13]. To address work performance issues, it is necessary to define and identify performance problems. Mitchell [14] suggests that work performance problems arise when workers encounter challenges that require physical or mental effort to resolve. The presence and types of performance problems, including physical and mental health conditions and vitality issues, can be assessed through careful observation and the use of structured tools [14, 15].

Although various scales/tools have been used to assess each dimension of individual work performance, the Individual Work Performance Questionnaire (IWPQ) was specifically developed to measure all three dimensions comprehensively [2]. The questionnaire consists of 18 items divided into three subscales, each assessing one of the dimensions. Originally developed in Dutch to review occupational health, psychology, and management literature and updated to the latest version (version 1.0) [16–18], the IWPQ scales were cross-culturally adapted into several languages, including English, Swedish, Spanish, Argentine Spanish, and Indonesian [19–23]. These studies highlight the IWPQ's versatility and its ability to provide accurate assessments of work performance across different regions and industries. The IWPQ's

structure, which incorporates both positive performance indicators (task and contextual performance) and negative indicators (counterproductive work behavior), allows for a comprehensive evaluation of an employee's contributions to the organization. This holistic approach is increasingly important in globalized work environments, where diverse cultural backgrounds can influence perceptions and evaluations of work performance [24].

Cultural factors exert a substantial influence on job performance, impacting employees' values, actions, communication methods, and overall work approach. Effective management of performance in a globalized workforce necessitates a comprehensive grasp of employees' cultural background [25]. Divergent cultural norms might result in different anticipations regarding leadership, decision-making, and collaboration, potentially either augmenting or impeding work effectiveness based on the adeptness of managing these disparities. In collectivist societies, there is a greater emphasis on teamwork and maintaining group peace, while individualist cultures place a higher importance on personal achievements [26]. These cultural subtleties can have an influence on motivation, job contentment, and ultimately, work productivity. Moreover, cultural perspectives on authority, unpredictability, and work-life balance can have a substantial impact on employee performance and workplace dynamics [27]. The IWPQ is a questionnaire utilized to evaluate work performance in many cultural settings. The measurement encompasses three primary dimensions: task performance, contextual performance, and counterproductive work behavior. Task performance pertains to the proficiency and efficacy in carrying out job responsibilities, contextual performance encompasses behaviors that contribute to the overall organizational atmosphere, and counterproductive work behavior encompasses activities that inflict harm upon the organization or its members [28]. Cultural factors can impact the perception and assessment of these aspects. For instance, in a society that greatly esteems communal achievement, contextual performance may be evaluated more positively, whereas in cultures that prioritize individual accomplishment, task performance may be given greater importance [29]. Gaining a comprehensive understanding of these cultural influences is essential for appropriately interpreting the results of the IWPQ and for effectively applying performance management techniques in various cultural contexts. The IWPQ scales were translated into and validated in different languages, and their English versions were studied and found to be valid in South Africa [30]. Although the IWPQ has been adapted into several languages, it has not yet been translated into Turkish. The purpose of this work is to translate the IWPQ into Turkish, evaluate the validity and reliability of the Turkish translation, and offer a cross-culturally appropriate

instrument that will help Turkish academics share their findings and make cross-national comparisons easier. Consequently, it is crucial to verify whether the instrument functions well for workers of different sociodemographic characteristics.

Materials and methods

Study design and participants

This study employed a methodological approach to assess the validity and reliability of the Turkish version of the Individual Work Performance Questionnaire (IWPQ). The participants consisted of employees who had been working regularly for 8 h a day and 40 h a week were included in the research. Based on the guideline that recommends a minimum of 10 participants per item for validity and reliability analyses [31], a minimum sample size of 180 participants was targeted (as the IWPQ scales have 18 items). The inclusion criteria were as follows: (1) aged between 18 and 65 years, (2) working regularly for at least 8 h a day, and (3) being a native Turkish speaker. Individuals with any neurological and psychiatric diseases were excluded from the study.

Before the study, the approval of the University of Health Sciences Gülhane Scientific Research Ethics Committee was obtained (approval date: 06.17.2021, approval number: 2021/284). Furthermore, the study

was conducted in accordance with the Declaration of Helsinki. An online informed consent form was signed by the participants. Participants were reached using the snowball sampling method. Participants were recruited from local private companies located in Ankara, Turkey. They were invited to participate in the study using an invitation email. The data of the study were collected through an online questionnaire using Google forms. The participants completed the scales, which took approximately 10–15 min. A reminder email was also sent to eligible participants to encourage participation one week after the initial email.

Study population

This is a methodological study that was carried out with 340 employees. The data of all participants were analysed as the first group; furthermore, a second session was performed for retest analysis with 64 individuals who also participated in the first session. The demographics and socioeconomic characteristics of the participants are presented in Table 1. Nearly half of the study population was female, and their median age was 36.0 years. The majority of the participants graduated from a university or a postgraduate degree. Although only seven individuals had a manager/administrator occupation and 12 participants were in the blue-collar working class, the

Table 1 Demographics and socioeconomic characteristics of the participants

Characteristics		Whole study group n = 340 (%)	Retest group n = 64 (%)
Gender, n (%)	Female	181 (53.2)	23 (35.9)
	Male	159 (46.8)	41 (64.1)
Age (years), Median (IQR)		36.0 (29.0–47.0)	44.0 (34.3–54.8)
Educational attainment, n (%)	Primary school	5 (1.5)	0 (0.0)
	Middle school	4 (1.2)	2 (3.1)
	High school	29 (8.5)	9 (14.1)
	University	192 (56.5)	13 (20.3)
	Postgraduate	110 (32.4)	40 (62.5)
Occupation, n (%)	Manager	7 (2.1)	0 (0.0)
	White-collar	321 (94.4)	62 (96.9)
	Blue-collar	12 (3.5)	2 (3.1)
Job tenure, n (%)	3–6 months	41 (12.1)	4 (6.3)
	6 months–1 year	31 (9.1)	3 (4.7)
	2–3 years	58 (17.1)	7 (10.9)
	4–5 years	54 (15.9)	9 (14.1)
	6–10 years	65 (19.1)	17 (26.6)
	11 years or more	91 (26.8)	24 (37.5)
Perceived economic status, n (%)	Very low	14 (4.1)	3 (4.7)
	Low	23 (6.8)	2 (3.1)
	Middle-low	57 (16.8)	7 (10.9)
	Middle	175 (51.5)	41 (64.1)
	Middle-high	61 (17.9)	8 (12.5)
	High	7 (2.1)	2 (3.1)
	Very high	3 (0.9)	1 (1.6)
Presence of chronic conditions, n (%)		75 (22.1)	21 (32.8)

IQR: 25–75% interquartile range

vast majority of the study population was white-collar workers (94.4%). While approximately a quarter of the individuals had been working for 11 years or more, the rest of them showed a similar frequency of distribution under 10 years of tenure. The perceived economic status, which was evaluated by a Likert-type question, had a normal distribution pattern. Of the participants, 22.1% had at least a chronic health condition. While most of the characteristics of the individuals in the retest group were broadly similar to those of the whole study population, women comprised the majority (64.1%), and the median age was 44.0 years in the retest group (Table 1).

Instruments

The questionnaire included the demographics and socio-economic characteristics of the participants, the IWPQ scales, and the UWES-3 scale.

The IWPQ scales have three subdimensions: task performance (items 1–5), contextual performance (items 6–13), and counterproductive work behavior (CWB) (items 14, 15, 16, 17 and 18) in the English version [32]. The scales utilize a five-point Likert rating (0 is for “seldom”, 1 is for “sometimes”, 2 is for “regularly”, 3 is for “often” and 4 is for “always” for task performance scale and contextual performance scale; and 0 is for “never”, 1 is for “seldom”, 2 is for “sometimes”, 3 is for “regularly” and 4 is for “often” for CWB scale), and a recall period of 3 months. The mean score of each subscale was the main output (16, 33). The English version of the IWPQ scales was adapted to Turkish and then administered to 340 participants in the first session. Next, the same scale was readministered to 74 participants to analyse the test-retest reliability three months later.

The Utrecht Work Engagement Scale (UWES)-9, which is widely used to assess work engagement, can be used for comparison purposes to establish convergent validity with the IWPQ scales [33, 34]. The UWES-3, a brief unidimensional version of the UWES-9 scale, has demonstrated acceptability in assessing work engagement and predicting various work-related conditions, such as job satisfaction, stress overload, perceived efficacy, and minor accidents [35]. Guler et al. [36] adapted the UWES-3 into Turkish and confirmed its validity and reliability. The UWES-9 was developed by Schaufeli et al. [37] as a shortened form of the 17-item version. It was used to evaluate the convergent validity of the IWPQ scales (34, 35). UWES-3, on the other hand, is the ultra-short and unidimensional version of UWES-9, which has a three-dimensional construct [38]. In our study, the UWES-3 was preferred as a comparison scale for convergent validity.

Adaptation procedure

The Turkish adaptation of the IWPQ scales followed a cross-cultural adaptation procedure based on the guidelines developed by Capik et al. [39]. First, the English version of the IWPQ scales was translated into Turkish by two native English speakers who were fluent in Turkish. The authors (EÖ and SBÇ) conducted a comparison of the two translations and rectified any discrepancies, resulting in the creation of a preliminary Turkish version. The specialists, all of whom held a PhD and specialized in rehabilitation, examined the text presented in a well-organized document that included both the original and Turkish copies placed next to each other. Their feedback was solicited to acquire expert perspectives. Typically, specialists provided their viewpoints on the representation of objects in the frequently employed societal format. The comments were assessed and a Turkish version was composed by the authors (EÖ and SBÇ). A consensus was reached to create a joint Turkish version. Another two native English speakers, proficient in Turkish, independently translated the initial translation of the document back into English. Both translators were unaware of the research objectives and had no prior knowledge of the English version of the IWPQ. The two back translations exhibited linguistic comparability between the original scale and the translated version, as they were similar both in terms of language and meaning. Any inconsistencies or uncertainties were addressed and resolved by the authors (EÖ and SBÇ). A pilot study was conducted with 25 individuals to assess the comprehensibility of the Turkish version, and feedback from the pilot study confirmed the understandability of all items. Following the pilot research, the Turkish version of the scale was modified slightly to incorporate the suggestions provided by the participants, resulting in the finalization of the Turkish IWPQ. The final version was sent to the scale's creator (LK) for approval, finalizing the adaptation procedure.

Statistical analysis

Statistical analyses were performed using SPSS version 20 software (IBM Corp. in Armonk, NY). The descriptive statistics of the study variables are presented as the mean with standard deviation (SD) for numerical data and as the frequency (n) with percentage (%) for nominal/ordinal data. The psychometric properties of the Turkish version of the IWPQ scales were assessed through validity and reliability analyses. Initially, exploratory factor analysis (EFA) using principal component analysis (PCA) with Promax rotation with Kaiser normalization (Kappa=4) was conducted to assess the factor structure of the scale. The Kaiser–Meyer–Olkin test was used to evaluate the adequacy of the sample. Bartlett's test of sphericity was performed to evaluate the level of the

correlations between the items. Cronbach's alpha score was used to assess internal consistency. A range between 0.700 and 0.900 was accepted as the reference range for Cronbach's alpha coefficient [40]. The intraclass correlation coefficient (ICC) with a 95% confidence interval (CI) between the first administration of the IWPQ scales to the whole study group and the second administration to the retest group was used to assess the test-retest reliability [41]. Spearman correlation coefficients between the IWPQ scales and the UWES-3 scale were used to analyse convergent validity. The acceptable strength of the correlation between the scales was defined as the coefficient being higher than 0.800 [42]. Interitem Spearman correlation coefficients were used to assess the internal consistency reliability. Furthermore, the average interitem correlation was calculated, and a range between 0.200 and 0.400 was defined as acceptable [43]. The histogram of the IWPQ scale scores of the whole study population and the frequency of the lowest or highest possible scores were used to assess the floor and ceiling effect. The highest limit for those scores was accepted as 15% for floor and ceiling effect criteria [44]. A *p* value less than 0.05 was defined as statistically significant.

Results

Validity and reliability analysis

Bartlett's test of sphericity was statistically significant ($\chi^2=3107.609$ and $p<0.001$), and the KMO coefficient was 0.885. According to these results, the sample size of the study was approved as adequate for PCA. Initially, PCA was performed without using a rotation method, and a three-factor structure with a cumulative variance of 60.7% was revealed (Table 2; Fig. 1). Furthermore, the items in the factors that were revealed in our model overlapped with the items in the original English version. However, several cross-loadings (items 1, 2, 3, 4, 5, 11 and 12) were found at that stage (data not shown). Next, PCA was reperformed using Varimax, Quartimax, Equamax, Direct Oblimin, and Promax rotation methods to minimize the number of cross-loaded items [45]. The number of cross-loaded items was reduced to one item (item 12), which was the least, using Promax rotation. Although the data were reanalyzed using other extraction methods, reducing factor numbers, and removing cross-loaded items, these approaches did not contribute to the scale construct and did not provide any explanation for variance. Finally, the three-factor structure of the scale was decided to report using original names: task performance (items 6–13), contextual performance (items 1–5), and counterproductive work behavior (items 14–18) (Table 2).

Table 2 Principal component analysis results

Item no	Factor structure and factor loadings			Total
	Factor 1 (Contextual performance)	Factor 2 (Task performance)	Factor 3 (Counterproductive work behavior)	
Item 1	-0.142	0.891	0.010	60.7%
Item 2	-0.025	0.833	0.080	
Item 3	-0.017	0.821	0.028	
Item 4	0.051	0.779	-0.042	
Item 5	0.060	0.751	-0.086	
Item 6	0.564	0.264	0.006	
Item 7	0.833	-0.007	0.023	
Item 8	0.559	0.294	0.003	
Item 9	0.586	0.286	0.001	
Item 10	0.699	0.149	-0.039	
Item 11	0.887	-0.105	0.023	
Item 12	0.920	-0.329	-0.004	
Item 13	0.664	0.039	-0.003	
Item 14	0.108	-0.032	0.714	
Item 15	0.049	0.018	0.783	
Item 16	-0.041	-0.002	0.697	
Item 17	-0.087	0.011	0.755	
Item 18	-0.013	0.021	0.771	
EVR	36.0%	15.9%	8.8%	
Eigenvalue	6.48	2.86	1.59	
Cronbach's alpha	0.894	0.875	0.796	

EVR: Explained variance ratio

Note Promax rotation with Kaiser normalization (Kappa=4) was used in PCA

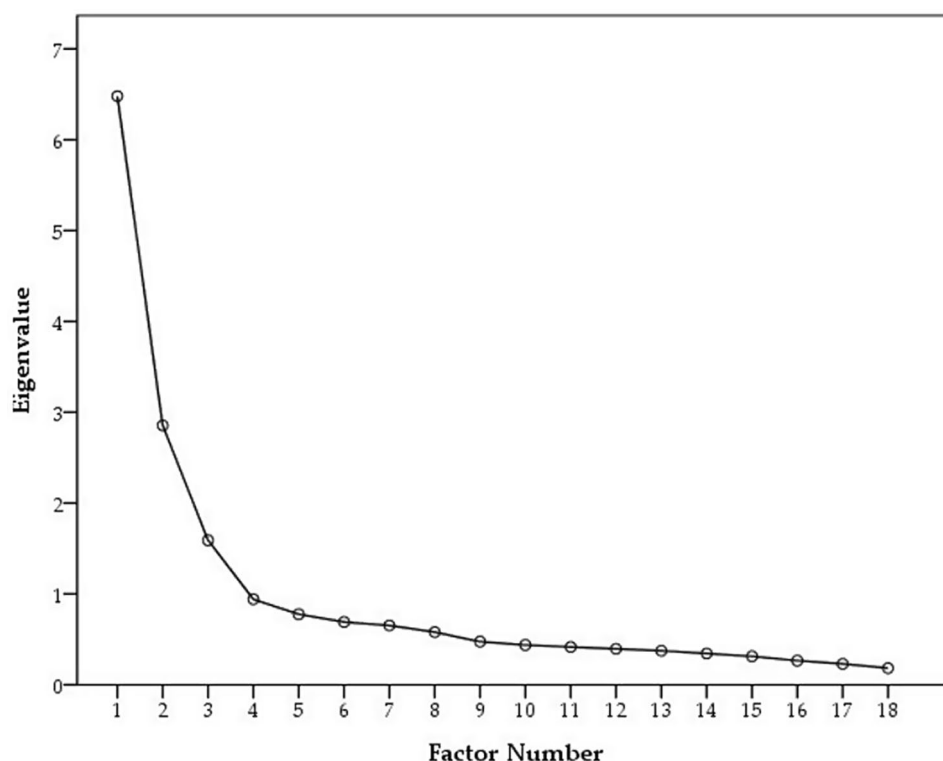


Fig. 1 Scree plot of exploratory factor analysis of the IWPQ scale. The figure shows a decrease from an Eigenvalue of 6.48 to an Eigenvalue of 1 in three phases

Cronbach's alphas of three subscales (task performance, contextual performance, and counterproductive work behavior) were 0.894, 0.875, and 0.796, respectively. For the whole scale, on the other hand, a Cronbach's alpha of 0.854 was found for the individuals in the first session group (Table 2).

The descriptive statistics and ICCs of the IWPQ scales for the whole study population and retest sample group are summarized in Table 3. The mean score of each item and the mean total score of all IWPQ scales of the whole study population were broadly similar to the means of the retest group. The ICCs (95% CIs) of task performance, contextual performance and CWB subscales between the whole group and retest group were 0.96 (0.94–0.98), 0.98 (0.97–0.99), and 0.91 (0.86–0.95), respectively (Table 3).

The correlations between IWPQ scales and UWES-3 are shown in Table 4 and scatter plots in Fig. 2. There was a statistically significant weak positive correlation between UWES-3 and the task performance subscale ($\rho=0.367$ and $p<0.001$), a statistically significant medium positive correlation between UWES-3 and the contextual performance subscale ($\rho=0.403$ and $p<0.001$), and a statistically significant weak negative correlation between UWES-3 and the CWB subscale ($\rho=-0.352$ and $p<0.001$).

The interitem correlations of the IWPQ scale of the whole study population are summarized in Table 5. The average interitem correlation coefficient of the scale was

0.282. This result was acceptable in terms of the pre-defined range of 0.200 to 0.400 (see supplementary file).

The histogram of the IWPQ scale scores for the whole study group is shown in Fig. 3. Of the participants, 2.1% had the lowest possible score (0) on the task performance scale, 1.8% on the contextual performance scale, and 1.8% on the CWB scale. On the other hand, 4.4% of them achieved the highest score (4) on the task performance scale, 2.4% on the contextual performance scale, and 1.5% on the CWB scale. All these frequencies were acceptable in terms of floor and ceiling effects.

Discussion

In this methodological study, It was examined the reliability and validity of a self-reported Turkish version of the IWPQ that is short, novel, and intended for use in employees. In conclusion, our results indicate that the Turkish version of the IWPQ, which has a three-factor structure, is a reliable and valid instrument for assessing the work performance of employees. The scale's structural validity was verified by our analysis. The Turkish IWPQ is reliable, as evidenced by the test-retest findings and Cronbach's alpha.

Koopmans [46] reported that the perception of the items and the average scores between different occupational sectors (white/pink/blue) could be differential; furthermore, this was currently a question that has to

Table 3 Descriptive statistics and intraclass correlation coefficient between the first and second administrations of the IWPQ

Item no	Mean $\pm$ SD	
	Whole study group (<i>n</i> = 340)	Retest group (<i>n</i> = 64)
Item 1	2.31 $\pm$ 1.13	2.58 $\pm$ 1.17
Item 2	2.38 $\pm$ 1.10	2.38 $\pm$ 1.15
Item 3	2.30 $\pm$ 1.13	2.56 $\pm$ 1.08
Item 4	2.37 $\pm$ 1.11	2.44 $\pm$ 1.11
Item 5	2.08 $\pm$ 1.14	2.33 $\pm$ 1.07
Task performance	2.29 $\pm$ 0.92	2.45 $\pm$ 0.84
ICC (95% CI)	0.96 (0.94–0.98)	
Item 6	2.06 $\pm$ 1.30	2.05 $\pm$ 1.27
Item 7	2.13 $\pm$ 1.30	2.08 $\pm$ 1.34
Item 8	2.56 $\pm$ 1.16	2.58 $\pm$ 1.04
Item 9	2.66 $\pm$ 1.10	2.52 $\pm$ 1.05
Item 10	2.37 $\pm$ 1.23	2.33 $\pm$ 1.20
Item 11	2.06 $\pm$ 1.35	2.06 $\pm$ 1.30
Item 12	1.43 $\pm$ 1.27	1.48 $\pm$ 1.25
Item 13	2.34 $\pm$ 1.36	2.36 $\pm$ 1.31
Contextual performance	2.20 $\pm$ 0.95	2.18 $\pm$ 0.90
ICC (95% CI)	0.98 (0.97–0.99)	
Item 14	1.56 $\pm$ 1.06	1.16 $\pm$ 0.95
Item 15	0.75 $\pm$ 0.91	0.55 $\pm$ 0.71
Item 16	1.06 $\pm$ 1.04	0.78 $\pm$ 0.79
Item 17	1.79 $\pm$ 1.10	1.39 $\pm$ 0.95
Item 18	1.36 $\pm$ 1.12	1.02 $\pm$ 0.97
CWB	1.30 $\pm$ 0.78	0.98 $\pm$ 0.56
ICC (95% CI)	0.91 (0.86–0.95)	

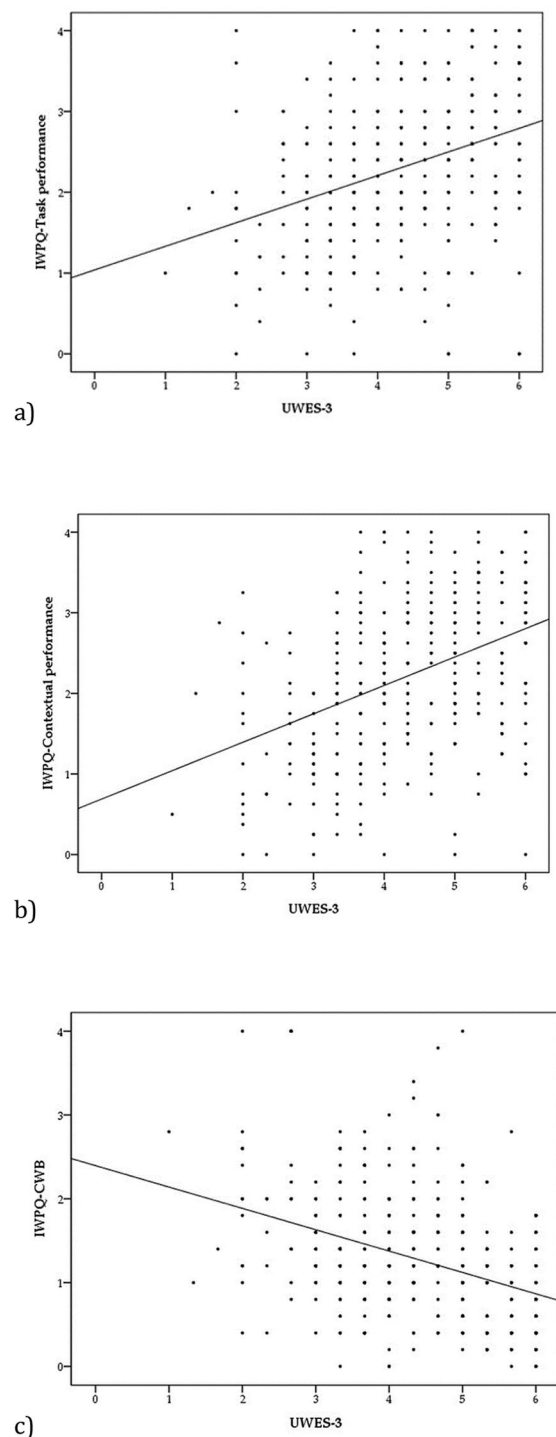
SD: standard deviation, ICC: intraclass coefficient between the whole group and retest group

Table 4 Correlations between IWPQ scales and UWES-3

		UWES-3
IWPQ-Task performance	ρ	0.367
	<i>p</i>	< 0.001**
	<i>n</i>	340
IWPQ-Contextual performance	ρ	0.403
	<i>p</i>	< 0.001**
	<i>n</i>	340
IWPQ-CWB	ρ	-0.352
	<i>p</i>	< 0.001**
	<i>n</i>	340

* $p < 0.05$ and ** $p < 0.001$

be solved. Beyond this point, we found that the Turkish version of the IWPQ scales has a three-factor structure, and all statistical validity and reliability measures were plausible and acceptable. There were several similarities and differences between our results and other adaptation studies on the IWPQ scales in the literature. However, Ployhart et al. [47] reported that cross-culturally adapted job performance scales demonstrated acceptable measurement invariance between employees whose mother

**Fig. 2** Scatter plots of the scores of (a) IWPQ-task performance scale, (b) IWPQ-contextual performance scale, (c) IWPQ-CWB scale and UWES-3 scale

tongue was different even when they were working in the same organizational environment, whereas the error variances of the scales and the construct validity patterns were broadly specific to the target language and culture.

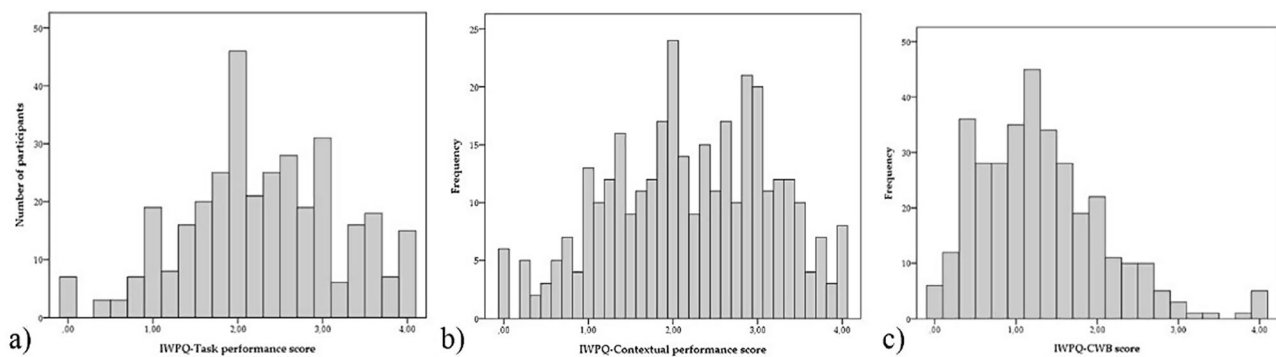


Fig. 3 Histogram of the scores of (a) IWPQ-task performance score, (b) IWPQ-contextual performance score, (c) IWPQ-CWB score

In most cross-cultural adaptation studies of the IWPQ scales, an 18-item and three-factor structure was revealed [20, 23, 32, 48]. The Argentinean version of the scale, on the other hand, showed a 13-item and three-factor structure [49]. However, it should be kept in mind that the participant demographics and occupational characteristics of all these studies were different, and all these features are closely related to work performance [50]. In our study, the three-factor structure and item dissemination into the factors overlapped with the original Dutch version and the English version, which we used as the source scale [2, 32, 51]. Furthermore, our Cronbach's alpha scores were comparable with other studies, in most of which alpha scores were reported as higher than 0.80 [20, 23, 32, 48, 49].

The Turkish version of the IWPQ scales showed acceptable internal consistency and test-retest reliability. Although all the studies in the literature analysed and reported their results on the internal consistency of the scale [20, 23, 32, 48, 49], none of them studied the test-retest reliability. One unique aspect of this study was the inclusion of test-retest reliability analysis, which is a valuable measure to assess the stability of the instrument over time. The high intraclass correlation coefficients (ICCs) obtained between the first session and retest session demonstrated strong test-retest reliability for the IWPQ scales.

The multitrait-multimethod matrix (MTMMM) concept was developed to analyse the convergent validity of different scales according to their similarity in terms of the construct (heterotrait or monotrait) and data collection (heteromethod or monomethod) [52]. It is expected that the strength of the correlation will increase from heterotrait-heteromethod comparison to monotrait-monomethod comparison [42]. In the present study, heterotrait-monomethod comparisons between the IWPQ scales (work performance) and UWES-3 (work engagement) were performed, and the results of these comparisons showed that the Turkish IWPQ scales had acceptable convergent validities in terms of

the MTMMM method [42, 52]. Koopmans et al. [33] reported similar results and argued that moderate to low positive correlations between work engagement and task and contextual performance and moderate to low negative correlations between work engagement and CWB were expectable and plausible.

Interitem correlations, which reflect the relationships between the items of a scale, are essential measures in validity studies. Analysing interitem correlations enables researchers to determine possible redundant items that assess the same content [53]. The average interitem correlation on a scale should be between 0.20 and 0.40 [43]. An average interitem correlation lower than 0.20 means that the items do not contain the same content dimension, and an average correlation higher than 0.40 means that the items do not explain the whole bandwidth of the scale structure. From this point, the balance between resemblance and diversity is an essential factor when analysing interitem correlations [43, 54]. In the present study, we found an average interitem correlation of 0.282, which indicated appropriate internal consistency reliability with an acceptable balance between the diversity and resemblance of the items. In the Swedish study, on the other hand, the average interitem correlation was 0.36 and comparable with our results [20].

In our study, all three scales had acceptable floor and ceiling effects. Floor and ceiling effects are quality measures in validity studies, and they are both expected to be lower than 15% [44]. There is a crucial balance between content validity and reliability in terms of participants' choice of extreme scores. Although the absence of extreme ends might indicate the lack of range variability in participants' responses, which negatively affects content validity, reliability may be negatively affected because of the lack of variance when participants do not choose extreme scores sufficiently [55, 56]. However, in the Swedish validation study, the authors reported that their CWB scale showed floor effects, and task performance and contextual performance scales showed ceiling effects [20]. Koopmans et al. [33] reported that floor and

ceiling effects could cause low responsiveness; people tended to score relatively high on the task performance scale and low on the CWB scale, and consequently, the floor and ceiling effects make it difficult to detect further improvement in task performance scale and possible decrease in CWB scale.

Despite the strengths of this study, including the comprehensive analysis of validity and reliability measures, there are some limitations to consider. First, the study sample predominantly consisted of white-collar workers, which may limit the generalizability of the findings to other occupational groups. The online data collection method may have also introduced sampling and non-sampling biases. However, online data collection more easily reached more people in the COVID-19 era. While the relatively small sample size of our study was sufficient for the statistical analysis, it may limit the external validity of the study results. Readers should keep all these limitations in their mind when interpreting the findings of the study. The Turkish validity and reliability study of the IWPQ significantly enhances the existing knowledge about the IWPQ by demonstrating its applicability across different cultural and linguistic contexts. By validating the scale in Turkish, the study confirms the robustness of the IWPQ's structure and its relevance in measuring work performance outside of its original context, thereby broadening its utility in international research. Additionally, the study offers insights into the linguistic and conceptual nuances that may arise when adapting performance measures for non-English-speaking populations, highlighting the importance of cultural sensitivity in psychometric evaluations. This research not only solidifies the IWPQ's credibility in the Turkish workforce but also sets a precedent for further cross-cultural adaptations, which are essential for comprehensive global work performance research. The Turkish validity and reliability study of the IWPQ offers several practical recommendations for its application in workplace settings. Firstly, the Turkish IWPQ is a valid and effective tool for assessing the work performance of employees. Secondly, the Turkish IWPQ can be utilized to identify areas of work performance that need to be addressed in employees and to measure the outcomes of interventions aimed at improving work performance. Additionally, the study highlights the importance of involving local employees in the adaptation process, suggesting that organizations should consider ongoing feedback from employees. Finally, the Turkish IWPQ can be used in comparative studies with other cultural adaptations of the scale, providing valuable data for multinational companies seeking to understand performance trends across different regions. This approach not only supports more comprehensive interventions but also contributes to work performance,

which constitutes an important area of activity in daily life.

Conclusions

According to the results of this study, a three-factor structure was revealed in factor analysis, and it was similar to the English version of the scale. Furthermore, other tests for validity and reliability presented acceptable and plausible findings. In conclusion, the Turkish version of the IWPQ scales is valid and reliable for evaluating the work performance of employees. However, further similar studies with participants with different characteristics, such as job tenure and occupation types, should be conducted to support and/or generalize our results.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40359-024-02010-2>.

Supplementary material 1: Table 5. Interitem correlation matrix of the IWPQ scale

Author contributions

EÖ planned the conceptualization and design of the study. EÖ completed the ethics committee process of the research. EÖ and SBÇ collected and analyzed the data and created the text of the article. EÖ, SBÇ and LK worked together to develop the article into its final version.

Funding

The authors declare that no funds, grants, or other support was received during the preparation of this manuscript.

Data availability

The data of the study are available upon reasonable request. The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Declarations

Ethical approval

Before the study, the approval of the University of Health Sciences Gülhane Scientific Research Ethics Committee was obtained (approval date: 06.17.2021, approval number: 2021/284). Furthermore, the study was conducted in accordance with the Declaration of Helsinki.

Consent to participate

Informed consent was obtained from all participants included in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 2 October 2023 / Accepted: 17 September 2024

Published online: 18 October 2024

References

1. Koopmans L, Bernaards CM, Hildebrandt VH, Schaufeli WB, de Vet HCW, van der Beek AJ. Conceptual frameworks of individual work performance: a systematic review. *J Occup Environ Med*. 2011;53(8):856–66.

2. Koopmans L, Bernaards CM, Hildebrandt VH, De Vet HC, Van Der Beek AJ. Construct validity of the individual work performance questionnaire. *J Occup Environ Med*. 2014;56(3):331–7.
3. Harrison DA, Newman DA, Roth PL. How important are Job attitudes? Meta-Analytic Comparisons of Integrative Behavioral Outcomes and time sequences. *Acad Manag J*. 2006;49(2):305–25.
4. Word D, Fredolin J, P, & Ballesteros, J. V. A. The effect of work ethics of employees on their work performance. *Research in Management and Humanities DWJMH*. 2022; 1(1): 58–82.
5. den Kamp EMO, Bakker AB, Tims M, Demerouti E, de Wijs JJ. Working with a chronic health condition: the implications of proactive vitality management for occupational health and performance. *J Vocat Behav*. 2024;150:103987.
6. Morales-García WC, Vallejos M, Sairitupa-Sanchez LZ, Morales-García SB, Rivera-Lozada O, Morales-García M. Depression, professional self-efficacy, and job performance as predictors of life satisfaction: the mediating role of work engagement in nurses. *Front Public Health*. 2024;12:1268336.
7. Gutiérrez OI, Polo JD, Zambrano MJ, Molina DC. Meta-analysis and scientific mapping of well-being and job performance. *Span J Psychol*. 2020;23:e43.
8. Schaufeli WB, Bakker AB, Van Rhenen W. How changes in job demands and resources predict burnout, work engagement, and sickness absenteeism. *J Organizational Behavior: Int J Industrial Occup Organizational Psychol Behav*. 2009;30(7):893–917.
9. Ganster DC, Rosen CC. Work stress and employee health: a multidisciplinary review. *J Manag*. 2013;39(5):1085–122.
10. Demerouti E, Bakker AB, Nachreiner F, Schaufeli WB. The job demands-resources model of burnout. *J Appl Psychol*. 2001;86(3):499.
11. Goetzel RZ, Long SR, Ozminkowski RJ, Hawkins K, Wang S, Lynch W. Health, absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting US employers. *J Occup Environ Med*. 2004;46(4):398–412.
12. Burton WN, Schultz AB, Chen CY, Edington DW. The association of worker productivity and mental health: a review of the literature. *Int J Workplace Health Manage*. 2008;1(2):78–94.
13. Anesetti-Rothermel A, Sambamoorthi U. Physical and mental illness burden: disability days among working adults. *Popul Health Manage*. 2011;14(5):223–30.
14. Mitchell JA. Analysing work performance problems—methods and their effectiveness. *Int J Rehabilitation Res Int Z fur Rehabilitationsforschung Revue Int De recherches de Readaptation*. 1995;18(1):27–34.
15. de Jonge J, Peeters MCW. The vital worker: towards sustainable performance at work. *Int J Environ Res Public Health*. 2019;16(6).
16. Koopmans L, Bernaards C, Hildebrandt V, Van Buuren S, Van der Beek AJ, De Vet HC. Development of an individual work performance questionnaire. *Int J Productivity Perform Manage*. 2012;62(1):6–28.
17. Koopmans L, Bernaards CM, Hildebrandt VH, Van Buuren S, Van der Beek AJ, De Vet H. Improving the individual work performance questionnaire using rasch analysis. *J Appl Meas*. 2014;15(2):160–75.
18. Koopmans L, Coffeng JK, Bernaards CM, Boot CR, Hildebrandt VH, de Vet HC, et al. Responsiveness of the individual work performance questionnaire. *BMC Public Health*. 2014;14:1–11.
19. Koopmans L, Bernaards CM, Hildebrandt VH, Lerner D, de Vet HC, van der Beek AJ. Cross-cultural adaptation of the individual work performance questionnaire. *Work*. 2016;53(3):609–19.
20. Dâderman AM, Ingelgård A, Koopmans L. Cross-cultural adaptation, from Dutch to Swedish language, of the individual work performance questionnaire. *Work (Reading Mass)*. 2020;65(1):97–109.
21. Ramos-Villagrasa PJ, Barrada JR, Fernández-del-Río E, Koopmans L. Assessing job performance using brief self-report scales: the case of the individual work performance questionnaire. *Revista De Psicología Del Trabajo Y De las Organizaciones*. 2019;35(3):195–205.
22. Gabini SM, Salessi SM. Validación de la escala de rendimiento laboral individual en trabajadores argentinos. 2016.
23. Widyastuti T, Hidayat R. Adaptation of individual work performance questionnaire (IWPO) into Bahasa Indonesia. *Int J Res Stud Psychol*. 2018;7(2):101–12.
24. Spector PE, Sanchez JI, Siu OL, Salgado J, Ma J. Eastern versus western control beliefs at work: an investigation of secondary control, socioinstrumental control, and work locus of control in China and the US. *Appl Psychol*. 2004;53(1):38–60.
25. Hofstede G. Cultural dimensions in management and planning. *Asia Pac J Manage*. 1984;1:81–99.
26. Tarique I, Schuler RS. Global talent management: literature review, integrative framework, and suggestions for further research. *J World Bus*. 2010;45(2):122–33.
27. Bakker AB, Demerouti E. The job demands-resources model: state of the art. *J Managerial Psychol*. 2007;22(3):309–28.
28. Koopmans L, Bernaards CM, Hildebrandt VH, Schaufeli WB, de Vet Henrica C, Van Der Beek AJ. Conceptual frameworks of individual work performance: a systematic review. *J Occup Environ Med*. 2011;53(8):856–66.
29. Caligiuri P, Tarique I. Dynamic cross-cultural competencies and global leadership effectiveness. *J World Bus*. 2012;47(4):612–22.
30. Vaart Lyd. The performance measurement conundrum: construct validity of the individual work performance questionnaire in South Africa. *The South African Journal of Economic and Management Sciences (SAJEMS)*. 2021;24(1):3581.
31. Kyriazos TA. Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General %J Psychology. 2018;Vol.09No.08:25.
32. Koopmans L, Bernaards CM, Hildebrandt VH, Lerner D, de Vet HC, van der Beek AJ. Cross-cultural adaptation of the individual work performance questionnaire. *Work*. 2016; 53(3):609–19.
33. Koopmans L, Coffeng JK, Bernaards CM, Boot CRL, Hildebrandt VH, de Vet HCW, et al. Responsiveness of the individual work performance questionnaire. *BMC Public Health*. 2014;14(1):513.
34. Santalla-Banderi Z, Alvarado JM. Structural and predictive validity of the Spanish short version of Utrecht Work Engagement Scale in Educational Institutions workers. *Span J Psychol*. 2022;25:e10.
35. Merino-Soto C, Lozano-Huamán M, Lima-Mendoza S, Calderón de la Cruz G, Juárez-García A, Toledano-Toledano F. Ultrashort Version of the Utrecht Work Engagement Scale (UWES-3): a psychometric Assessment. *Int J Environ Res Public Health*. 2022;19(2).
36. Güler M, Çetin F, Basim HN. The Validity and Reliability Study of Work Engagement Ultra Short Version (UWES-3): proposal of an Alternative Version (UWES-6). *J Hum Work*. 2019;6(2):189–97.
37. Schaufeli WB, Bakker AB, Salanova M. The measurement of Work Engagement with a short questionnaire: a cross-national study. *Educ Psychol Meas*. 2006;66:701–16.
38. Schaufeli WB, Shimazu A, Hakanen J, Salanova M, De Witte H. An Ultra-short measure for Work Engagement. *Eur J Psychol Assess*. 2017;35(4):577–91.
39. Çapık C, Gözüm S, Aksayan S. Intercultural scale adaptation stages, Language and Culture Adaptation: updated Guideline. *Florence Nightingale J Nurs*. 2018;26(3):199–210.
40. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011;2:53–5.
41. Koo TK, Li MY. A Guideline of selecting and reporting Intraclass correlation coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155–63.
42. Furr RM. Estimating and evaluating convergent and discriminant validity. In: Furr RM, editor. *Psychometrics: an introduction*. 3rd ed. Thousand Oaks, CA: Sage; 2018. pp. 253–307.
43. Piedmont RL. Inter-item correlations. In: Michalos AC, editor. *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht: Springer Netherlands; 2014. pp. 3303–4.
44. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Quality of life research. *Int J Qual life Aspects Treat care Rehabilitation*. 1995;4(4):293–307.
45. Brown J. Choosing the right type of Rotation in PCA and EFA. *Shiken: JALT Test Evaluation SIG Newsl*. 2009;13(3):20–5.
46. Koopmans L. Individual work performance questionnaire: instruction Manual. Amsterdam: TNO Innovation for Life - Vrije Universiteit University Medical Center; 2015.
47. Ployhart RE, Wiechmann D, Schmitt N, Sacco JM, Rogg K. The cross-cultural equivalence of job performance ratings. *Hum Perform*. 2003;16(1):49–79.
48. Ramos-Villagrasa PJ, Barrada JR, Fernández-del-Río E, Koopmans L. Assessing job performance using brief self-report scales: the case of the individual work performance questionnaire. *J Work Organizational Psychol*. 2019;35(3):195–205.
49. Gabini SM, Salessi SM. Validación de la escala de rendimiento laboral individual en trabajadores argentinos. *Evaluar*. 2016;16:31–45.
50. Campbell JP. Behavior, performance, and effectiveness in the twenty-first century. Kozlowski, Steve W. J. (ed.), *The Oxford handbook of organizational psychology*, Oxford library of psychology. Volume 1. New York, NY, US: Oxford University Press; 2012. pp. 159–94.

51. Koopmans L, Bernaards C, Hildebrandt V, van Buuren S, van der Beek AJ, de Vet HCW. Development of an individual work performance questionnaire. *Int J Productivity Perform Manage*. 2013;62(1):6–28.
52. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull*. 1959;56(2):81–105.
53. Cohen RJ, Swerdlik ME. *Psychological Testing and Assessment: an introduction to tests and measurement*. 7th ed. New York: McGraw-Hill; 2010.
54. Allen MJ, Yen WM. *Introduction to measurement theory*. Long Grove, Illinois: Waveland; 2002.
55. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42.
56. Garin O. Ceiling effect. In: Michalos AC, editor. *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht: Springer Netherlands; 2014. pp. 631–3.

Publisher's note

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