

Chapter 6

Construct Validity of the Individual Work Performance Questionnaire

Linda Koopmans, Claire M. Bernaards, Vincent H. Hildebrandt,
Henrica C.W. de Vet, Allard J. van der Beek

Journal of Occupational and Environmental Medicine. 2014; 56(3)

DOI 10.1097/JOM.000000000000113

6

Abstract

Objective: It is difficult to measure individual work performance comprehensively and generically. This study examines the construct validity of the recently developed Individual Work Performance Questionnaire (IWPQ).

Methods: 1,424 Dutch workers from three occupational sectors (blue, pink, and white collar) participated in the study. First, IWPQ scores were correlated with related constructs (convergent validity). Second, differences between known groups were tested (discriminative validity).

Results: First, IWPQ scores correlated weakly to moderately with absolute and relative presenteeism, and work engagement. Second, significant differences in IWPQ scores were observed for workers differing in job satisfaction, and workers differing in health.

Conclusion: Overall, the results indicate acceptable construct validity of the IWPQ. Researchers are provided with a reliable and valid instrument to measure IWP comprehensively and generically, amongst workers from different occupational sectors, with and without health problems.

Introduction

Individual work performance (IWP), defined as “*behaviors or actions that are relevant to the goals of the organization*” [1], is an important outcome in multiple research fields, as well as in practice. The conceptualization of IWP has a long history, and many frameworks have been proposed to describe the construct domain of IWP [e.g., 1-3]. The measurement of IWP has proven to be even more challenging, with numerous and diverse behaviors, actions, or results being applied as indicators of IWP [4]. Thus, despite the importance of IWP in research and practice, there is little consensus on how to conceptualize and measure IWP. This lack of consensus is undesirable, because valid measurement is a prerequisite for accurately establishing, for example, predictors of IWP, or effectiveness of interventions to improve IWP.

In the field of occupational health, there has been little attention for conceptualizing the IWP construct. The main focus was on sickness absenteeism or presenteeism, i.e., work absence or losses in IWP due to health impairments. In accordance, numerous instruments have been developed to measure sickness absenteeism or presenteeism, such as the Work Productivity And Impairment Questionnaire [5], Work Limitations Questionnaire [6], and the WHO Health and Performance Questionnaire [7].

In the field of work and organizational psychology, traditionally, the main focus of the IWP construct was on task performance, which can be defined as the proficiency with which individuals perform the core substantive or technical tasks central to his or her job [1]. It is now generally agreed upon that, in addition to task performance, the IWP domain consists of contextual performance and counterproductive work behavior [2,3,8]. Contextual performance can be defined as behaviors that support the organizational, social and psychological environment in which the technical core must function [9]. Counterproductive work behavior can be defined as behavior that harms the well-being of the organization [3]. In accordance, work and organizational psychologists have developed numerous scales to measure task performance [e.g., 10], contextual performance [e.g., 11], or counterproductive work behavior [e.g., 12].

It is evident that a multitude of instruments exists to measure IWP, or related constructs such as absenteeism or presenteeism. However, these existing scales show several limitations. Most strikingly, none of them measure all of the relevant dimensions of IWP together. Thus, they do not measure the full range of IWP. Also, scales measuring different dimensions can include antithetical items,

creating unjust overlap between these scales [13]. As a result, the content validity of these scales can be questioned. Furthermore, none of the scales appear suitable for generic use. The scales were developed for specific populations, such as employees with health problems [e.g., 5-7,13], or they were developed and refined based on employees with a specific occupation [e.g., 10,11].

Recently, the Individual Work Performance Questionnaire 1.0 [IWPQ; 14,15] was developed, in order to overcome limitations of existing questionnaires. A conceptual framework for the IWPQ was established based on a systematic review of the occupational health, work and organizational psychology, and management and economics literature [8]. The conceptual framework consists of three dimensions (task performance, contextual performance, and counterproductive work behavior), that represent the full range of IWP. No antithetical items were included in the questionnaire. Furthermore, the IWPQ was developed and refined based on a generic working population (i.e., blue, pink, and white collar workers), and is therefore suitable across occupational sectors, as well as for workers with and without health problems. The development, as well as the face, and structural validity of the IWPQ have been established in previous studies [4,14,15].

The current study expands research on the IWPQ by examining its construct validity. Construct validity refers to whether the instrument provides the expected scores, based on existing knowledge about the construct [16]. In the current study, construct validity was assessed by testing expectations about the relationship of the IWPQ scales with related constructs (convergent validity), and about differences between groups (discriminative validity).

Convergent validity

First, the IWPQ was correlated with the World Health Organization's Health and Performance Questionnaire [HPQ; 7], a validated questionnaire that intends to measure a similar construct. Only the HPQ presenteeism questions were administered. If scores on these questionnaires correlate strongly, then this indicates convergent validity [16].

Hypothesis 1a: The HPQ absolute presenteeism score, representing one's overall performance, was expected to show a strong positive correlation with the IWPQ task and contextual scales, and a strong negative correlation with the IWPQ counterproductive scale ($r > 0.50$ and $r < -0.50$, respectively [17]).

Hypothesis 1b: The HPQ relative presenteeism score, representing one's overall performance compared to the performance of most workers at the same job, was expected to show a weak positive correlation with the IWPQ task and contextual scales ($r < 0.30$), and a weak negative correlation with the IWPQ counterproductive scale ($r > -0.30$).

Second, the IWPQ was correlated with the Utrecht Work Engagement Scale [UWES; 18]. Work engagement is defined as “*a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption*” [19]. We consider work engagement and individual work performance to be related but not similar constructs. Several studies have shown that work engagement is moderately positive related to IWP [e.g., 20,21]. If scores on these questionnaires correlate moderately, then this indicates convergent validity.

Hypothesis 1c: The overall work engagement score, and its subscale scores, were expected to show a moderate positive correlation with the IWPQ task and contextual scales, and moderate negative correlation with the IWPQ counterproductive scale ($r = 0.30-0.50$ and $r = -0.30- -0.50$, respectively).

Discriminative validity

Another way to test the validity of the IWPQ is to examine whether it can differentiate known groups. Two known predictors of IWP are job satisfaction [e.g., 22] and health [e.g., 23]. Job satisfaction can be defined as “*an emotional state resulting from the evaluation or appraisal of one's job experiences*” [24]. A recent review [25] and meta-analysis [22] have shown that higher job satisfaction predicts higher IWP. Health can be defined as “*a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*” [26]. Both mental health [e.g., 27] and physical health [e.g., 23] have shown to be predictors of IWP.

Hypothesis 2a: Persons high in job satisfaction were expected to show significantly higher IWPQ task and contextual scores, and lower IWPQ counterproductive scores, than persons low in job satisfaction.

Hypothesis 2b: Persons in good health were expected to show significantly higher IWPQ task and contextual scores, and lower IWPQ counterproductive scores, than persons in poor health.

Methods

Participants

Participants were selected from a representative sample of Dutch workers from three occupational sectors: blue collar (manual workers, e.g.,: carpenter, mechanic, truck driver), pink collar (service workers, e.g.,: hairdresser, nurse, teacher), and white collar workers (office workers, e.g.,: manager, architect, scientist). Participants were recruited from a large internet panel consisting of Dutch adults willing to participate in research projects in exchange for a small reward.

Measures

Individual work performance was measured using the Individual Work Performance Questionnaire 1.0 [15]. The IWPQ 1.0 consisted of 3 scales (task performance, contextual performance, and counterproductive work behavior) with a total of 18 items (see Table 1). Within each scale, items were presented to participants in randomized order, to avoid order effects. All items had a recall period of 3 months and a 5-point rating scale (“*seldom*” to “*always*” for task and contextual performance, “*never*” to “*often*” for counterproductive work behavior). For the IWPQ subscales, a mean score was calculated by adding the item scores, and dividing their sum by the number of items in the subscale. Hence, the IWPQ yielded three subscale scores that ranged between 0 and 4, with higher scores reflecting higher task and contextual performance, and higher counterproductive work behavior.

To examine convergent validity, two presenteeism questions of the HPQ [7] were used: “*On a scale from 0 to 10 where 0 is the worst job performance anyone could have at your job and 10 is the performance of a top worker, how would you rate the usual performance of most workers in a job similar to yours*” (item 1) and “*Using the same 0-to-10 scale, how would you rate your overall job performance on the days you worked*” (item 2). Both had a recall period of 3 months and an 11-point rating scale (“*worst performance*” to “*top performance*”). The HPQ absolute presenteeism score was calculated by multiplying item 2 (*rating of overall job performance*) by 10. It has a lower bound of 0 (total lack of performance during time on the job) and an upper bound of 100 (no lack of performance during time on the job). The HPQ relative presenteeism score was calculated by dividing item 2 by item 1. It is restricted to the range of 0.25 to 2.00, where 0.25 is the worst relative performance (25% or less of other workers’ performance) and 2.00 is the best performance (200% or more of other workers’ performance [28]).

Work engagement was measured using the UWES-9 [18]. The UWES includes three scales (vigor, dedication, and absorption) and has been extensively validated [e.g., 19,20]. No recall period was specified and all items had a 7-point rating scale (“almost never” to “always”). The UWES-9 yielded one total score and three subscale scores ranging between 0 and 6.

Job satisfaction and overall health were measured using one question each (“How do you rate your overall job satisfaction?” and “How do you rate your overall health?”). Both had a recall period of 3 months and an 11-point rating scale (“very low” to “very high”). Research has shown that a single-item measure of job satisfaction and health-related quality of life correlate highly with multi-item measures of job satisfaction and health-related quality of life, and can therefore be considered valid [e.g., 29-31].

Table 1. Items of the Individual Work Performance Questionnaire (IW PQ), with means and standard deviations (SD) on a 0-4 range

Items	Mean	SD
Task performance (TP) scale		
In the past 3 months...		
TP1	I managed to plan my work so that it was done on time.	2.80 0.95
TP2	My planning was optimal.	2.47 0.98
TP3	I kept in mind the results that I had to achieve in my work.	3.11 0.81
TP4	I was able to separate main issues from side issues at work.	2.83 0.82
TP5	I was able to perform my work well with minimal time and effort.	2.32 1.00
Contextual performance (CP) scale		
In the past 3 months...		
CP1	I took on extra responsibilities.	2.24 1.09
CP2	I started new tasks myself, when my old ones were finished.	2.57 1.13
CP3	I took on challenging work tasks, when available.	2.32 1.08
CP4	I worked at keeping my job knowledge up-to-date.	2.28 1.15
CP5	I worked at keeping my job skills up-to-date.	2.42 1.02
CP6	I came up with creative solutions to new problems.	2.31 0.98
CP7	I kept looking for new challenges in my job.	2.12 1.10
CP8	I actively participated in work meetings.	2.25 1.20

Table 1. Continued

Counterproductive work behavior (CWB) scale			
In the past 3 months...			
CWB1	I complained about unimportant matters at work.	0.97	0.85
CWB2	I made problems greater than they were at work.	0.71	0.76
CWB3	I focused on the negative aspects of a work situation, instead of on the positive aspects.	1.10	0.86
CWB4	I spoke with colleagues about the negative aspects of my work.	1.56	1.02
CWB5	I spoke with people from outside the organization about the negative aspects of my work.	1.21	1.05

Data analysis

Pearson's correlations of the IWPQ subscale scores with the HPQ absolute and relative presenteeism scores, and the UWES-9 scores, were calculated. In addition, the correlations were examined separately for each occupational sector (blue, pink, and white collar workers), to determine whether the correlation strengths differed across occupational sectors.

For discriminative validity, groups low and high in job satisfaction, and low and high in overall health, were created using quartiles. Analyses of variance (ANOVAs) were performed to examine whether the job satisfaction quartiles, and the overall health quartiles, significantly differed on the IWPQ scores. Post-hoc tests with Bonferroni correction were performed to determine which quartiles significantly differed from each other. Finally, the ANOVAs with Bonferroni correction were performed separately for each occupational sector (blue, pink, and white collar workers), to determine whether the findings were generalizable across occupational sectors. All analyses were performed in SPSS 20 [32].

Results

Participants

In January 2012, 1,424 Dutch workers filled in the IWPQ. Participants were all employed, and aged 17 to 69 years. Table 2 presents sample characteristics and mean (and SD) scores on the IWPQ scales.

Table 2. Sample characteristics, and mean and standard deviation (SD) scores on the IWPQ task performance, contextual performance, and counterproductive work behavior scales (range 0-4)

	Total sample (N = 1,424)	Blue collar (n = 442)	Pink collar (n = 540)	White collar (n = 442)
Gender (% female)	42.4	14.0	65.6	42.5
Age (%)				
17-34 years	22.2	19.9	23.0	23.5
35-44 years	26.2	22.6	30.7	24.2
45-54 years	29.6	29.4	28.9	30.5
55-69 years	22.0	28.1	17.4	21.8
Education level (%)				
Primary	3.1	5.4	3.3	0.5
Secondary	38.1	54.5	40.7	18.6
Middle-level applied	29.7	34.4	34.4	19.2
Higher professional	28.5	4.8	21.3	61.1
Unknown	0.6	0.9	0.2	0.7
	Total sample Mean (SD)	Blue collar Mean (SD)	Pink collar Mean (SD)	White collar Mean (SD)
Task performance	2.71 (0.66)	2.82 (0.65)	2.71 (0.66)	2.59 (0.63)
Contextual performance	2.31 (0.77)	2.30 (0.82)	2.31 (0.76)	2.34 (0.72)
Counterproductive work behavior	1.11 (0.67)	1.03 (0.63)	1.09 (0.71)	1.21 (0.66)

Convergent validity

The absolute presenteeism score showed a moderate positive correlation with the IWPQ task and contextual scales, and a weak negative correlation with the counterproductive scale (see Table 3). The relative presenteeism score showed a weak positive correlation with the IWPQ subscales. These patterns of correlations did not differ across occupational sectors.

The overall work engagement score showed a moderate positive correlation with the IWPQ task and contextual scales, and an almost moderate negative correlation with the counterproductive scale. The work engagement subscales scores showed a weak to moderate positive correlation with the IWPQ task and contextual scales, and a weak to moderate negative correlation with the counterproductive scale. Again, this pattern did not differ across occupational sectors.

Table 3. Convergent validity: Expected and observed correlations of the IWPQ task performance (TP), contextual performance (CP), and counterproductive work behavior (CWB) scales with the HPQ absolute and relative presenteeism scores, and work engagement scores

HPQ absolute presenteeism score	Expected correlation	Observed correlation
TP scale	Strong (> 0.50)	Moderate (0.39) ^
CP scale	Strong (> 0.50)	Moderate (0.33) ^
CWB scale	Strong (< -0.50)	Weak (-0.16) ^
HPQ relative presenteeism score		
TP scale	Weak (< 0.30)	Weak (0.09) *
CP scale	Weak (< 0.30)	Weak (0.11) *
CWB scale	Weak (> -0.30)	Weak (0.07) *
UWES scores		
TP scale	Moderate (0.30-0.50)	Weak to moderate (0.26-0.35) ^
CP scale	Moderate (0.30-0.50)	Moderate (0.40-0.43) *
CWB scale	Moderate (-0.30- -0.50)	Weak to moderate (-0.23- -0.31) ^

Note. * Expectation confirmed.

^ Expectation party confirmed.

Discriminative validity

Job satisfaction quartiles differed significantly on task performance, contextual performance, and counterproductive work behavior (see Figure 1). There was a clear dose-response relationship between job satisfaction and individual work performance. Persons high in job satisfaction showing higher task and contextual performance, and lower counterproductive work behavior, than persons low in job satisfaction. All quartiles significantly differed from each other.

Overall health quartiles differed significantly on task performance, contextual performance, and counterproductive work behavior (see Figure 1). Again, there was a clear dose-response relationship between overall health and individual work performance. Persons in good overall health showing higher task and contextual performance, and lower counterproductive work behavior, than persons in poor overall health. Almost all quartiles significantly differed from each other. Again, these patterns did not differ across occupational sectors.

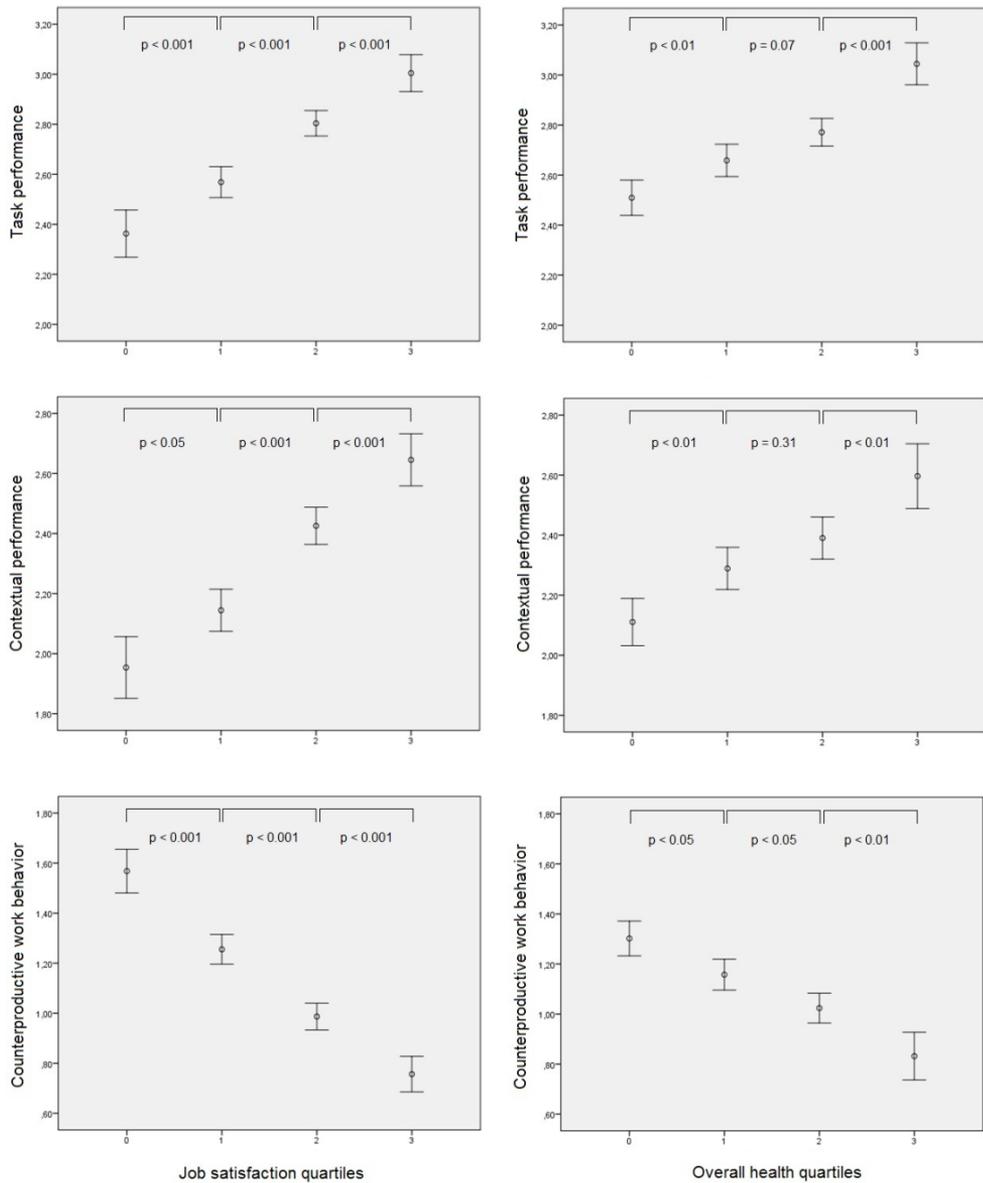


Figure 1. Discriminative validity of the IWPQ subscale scores (range 0-4) for job satisfaction quartiles and overall health quartiles (in the plots, the dot represents the mean score and the lines the 95% confidence interval).

Discussion

The current study expands research on the recently developed Individual Work Performance Questionnaire (IWPQ) [15] by examining its construct validity. Based on existing knowledge about the construct, the IWPQ largely provided the expected scores with regard to relationships to scores of other instruments, and with regard to differences between relevant groups. Terwee et al. [33] propose a standard of good construct validity when 75% of the hypotheses are confirmed. In the current study, 10 out of 15 (66.6%) hypotheses (5 constructs x 3 IWPQ scales) were confirmed. Overall, these findings indicate acceptable construct validity.

The correlations between the IWPQ subscale scores and the HPQ absolute presenteeism score were less strong than expected. Possibly, these weaker correlations were caused by limitations accompanying the administration of the HPQ in the current study. First, only the two presenteeism questions were administered in the current study. Although Kessler and colleagues [28] say that these questions can be abstracted from the complete HPQ, to our knowledge, the validity of these questions alone is unknown. The HPQ memory priming questions, intended to improve the accuracy of report, were omitted. Instead, the IWPQ questions were believed to be sufficient priming questions for participants to be able to give an overall rating of their work performance. Second, a longer recall period (3 months) was employed in the current study, instead of the original recall period (4 weeks). This was done to increase comparability between the questionnaires, and to avoid participants from needing to change their recall period while completing the questionnaires.

Another possible explanation for the weaker correlations is that the IWPQ concentrates on actual performance (what one typically does), whereas the HPQ absolute presenteeism question assesses actual performance in relation to possible performance (what one can maximally do). This subtle difference may make the two constructs less comparable than a priori expected, warranting moderate correlations. Furthermore, multi-item measures in which items measure the same construct are usually more reliable than single item measures [e.g., 31]. As IWP is a multi-dimensional construct, one overall question may not adequately reflect the full range of individual behaviors at work. The content that people have (or don't have) in mind when answering this question, may be substantially different for different persons, or from what researchers consider aspects of IWP. A finding to support this, is that the counterproductive work behavior scale correlated weakly with the HPQ

absolute presenteeism score. This suggests that people hardly take into account their counterproductive work behaviors, when rating their overall work performance. In the IWPQ, a more complete and generic picture of IWP is given, providing information about the separate components of IWP (a 'profile').

As expected, weak correlations of the IWPQ scores with the HPQ relative presenteeism score were found. This makes sense as the IWPQ does not ask workers to compare themselves with other workers, whereas this is the focus of the HPQ relative presenteeism score. Furthermore, participants may be uncomfortable rating the performance of their colleagues, or they may not have colleagues with similar jobs [34].

As expected, the IWPQ task and contextual performance scores showed moderate positive correlations with work engagement. The counterproductive work behavior score showed a weak to moderate negative correlation with work engagement. Although this last correlation was slightly lower than expected, the range of correlations was so close to expectations that this was not considered worrisome. The finding that engaged workers display more task and contextual performance behaviors, and less counterproductive work behaviors, is in line with previous studies which have shown that work engagement is positively related to IWP [e.g., 20,21]. Interestingly, the current study showed that work engagement was more strongly correlated with contextual behaviors, than with task or counterproductive behaviors. Thus, engaged workers mainly benefit their team or organization by engaging in behaviors that are not directly part of their central job tasks, but that do support the organizational, social, and psychological environment in which the central job tasks are performed.

Finally, the IWPQ was able to discriminate between relevant groups. Consistent with expectations, persons high in job satisfaction showed higher task and contextual scores, and lower counterproductive work behavior scores, than persons low in job satisfaction. Also, persons high in overall health showed higher IWPQ task and contextual scores, and lower IWPQ counterproductive work behavior scores, than persons low in overall health. The findings that satisfied workers, and healthy workers, perform better, is consistent with previous research [e.g., 22,23]. These findings also support the longstanding notion that happy, healthy workers are productive workers [e.g., 35].

Construct validity of related questionnaires

The construct validation process that we employed in the current study, appears to be similar to, or better than, the construct validation process of related questionnaires. Within the area of occupational health, several self-report questionnaires have been developed to measure losses in work performance (presenteeism or absenteeism). Contrary to the IWPQ, these questionnaires are mainly intended for persons with health complaints. The most used questionnaires in this area include the HPQ [7], Work Limitations Questionnaire [WLQ; 6] and Work Productivity and Activity Impairment questionnaire [WPAI; 5]. Although the construct validity of these questionnaires has been reasonably well established, a limitation of their construct validation process was that no a priori expectations were specified on the strengths of the correlations, and often, only the significance of a correlation was presented, and not the strength of a correlation.

Within work and organizational psychology, numerous self-report scales have been developed to measure task performance [e.g., 10], contextual performance [e.g., 11], or counterproductive work behavior [e.g., 12]. These scales are mainly used for establishing the determinants of IWP. Concerning their validity, the main focus has been on examining the relationships between the different scales [e.g., 36]. Relationships to scores of other instruments, or differences between relevant groups, hardly appear to have been examined. Thus, there is little information of the construct validity of these questionnaires.

Lack of a golden standard

A type of validity that could not be examined in the present study, is criterion validity, defined as “the degree to which scores of an instrument are an adequate reflection of a gold standard” [37]. This is because a perfect gold standard seldom exists in practice [16], as was the case for the construct under study here. A perfect gold standard may be an objective measure of individual work performance, however, these are very hard to obtain in practice [38]. Especially for knowledge work or high complexity jobs, direct measures of countable behaviors or outcomes such as production quantity or number of errors made, are almost impossible.

Although construct validity is often considered to be less powerful than criterion validation, with strong theories and specific and challenging expectations, it is possible to acquire substantial evidence that the measurement instrument is measuring what it purports to measure [16]. One of the strengths of the current study, is that a clear conceptual framework of IWP was present, and specific and

challenging expectations based on theory or literature findings were formulated. This makes it possible to draw firm conclusions about the construct validity of the IWPQ scores.

Future research

Construct validation is an ongoing process [16], and therefore, more research should be conducted to create a strong web of evidence to support the validity of the IWPQ. Strictly speaking, one can never state that a measurement instrument is valid, only that it provides valid scores in the specific situation in which it has been tested [16]. Future research will also need to examine additional properties of the IWPQ, such as its sensitivity to change, and interpretability of change scores. Also, it can be hypothesized that determinants of IWP affect the separate dimensions differently (as was shown in the current study for work engagement). Therefore, relationships of determinants to the separate dimensions of IWP can be further examined, as well as outcomes related to the separate dimensions of IWP. Other interesting avenues include relating IWPQ scores to objective outcomes (e.g., absenteeism data, monetary outcomes), or by comparing self-, other-, and supervisor-ratings (360 degrees performance rating).

Conclusion

Based on the results of the current study, the IWPQ 1.0 showed acceptable overall construct validity. Its convergent validity proved to be sufficient, and its discriminative validity very good. The IWPQ provides researchers with a reliable and valid instrument to measure IWP generically, amongst workers from different occupational sectors, and workers with and without health problems.

References

1. Campbell JP. Modeling the performance prediction problem in industrial and organizational psychology. In Dunnette MD, Hough LM, eds. *Handbook of industrial and organizational psychology, Vol.1 (2nd ed.)*. Palo Alto, CA, US: Consulting Psychologists Press; 1990:687-732.
2. Viswesvaran C, Ones DS. Perspectives on Models of Job Performance. *International Journal of Selection and Assessment*. 2000;8(4):216-226.
3. Rotundo M, Sackett PR. The relative importance of task, citizenship, and counterproductive performance to global ratings of performance: a policy-capturing approach. *J Appl Psychol*. 2002;87(1):66-80.
4. Koopmans L, Bernaards CM, Hildebrandt VH, De Vet HCW, van der Beek AJ. Measuring Individual Work Performance: Identifying and Selecting Indicators. *Work: Journal of Prevention, Assessment & Rehabilitation*. 2013;45(3).
5. Reilly MC, Zbrozek AS, Dukes EM. The Validity and Reproducibility of a Work Productivity and Activity Impairment Instrument. *Pharmacoeconomics*. 1993;4(5).
6. Lerner D, Amick BC, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. *Medical Care*. 2001;39(1):72-85.
7. Kessler RC, Barber C, Beck A, Berglund P, Cleary PD, McKenas D, et al. The World Health Organization Health and Work Performance Questionnaire (HPQ). *Journal of Occupational and Environmental Medicine*. 2003;45:156-174.
8. Koopmans L, Bernaards CB, Hildebrandt VH, Schaufeli WB, De Vet HCW, Van der Beek AJ. Conceptual frameworks of individual work performance: a systematic review. *Journal of Occupational and Environmental Medicine*. 2011;53(8):856-866.
9. Borman WC, Motowidlo SJ. Expanding the criterion domain to include elements of contextual performance. In: Schmitt N, Borman WC, eds. *Personnel Selection in Organizations*. San Francisco, CA: Jossey Bass; 1993. p. 71-98.
10. Williams LJ, Anderson SE. Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*. 1991;17(3):601-617.
11. Podsakoff PM, MacKenzie SB. *A second generation measure of organizational citizenship behavior*. Indiana University, Bloomington; 1989.

12. Bennett RJ, Robinson SL. Development of a Measure of Workplace Deviance. *J Appl Psychol.* 2000;85(3):349-360.
13. Dalal RS. A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *J Appl Psychol.* 2005;90:1241-1255.
14. Koopmans L, Bernaards CM, Hildebrandt VH, Van Buuren S, Van der Beek AJ, De Vet HCW. Development of an Individual Work Performance Questionnaire. *International Journal of Productivity and Performance Management.* 2013;62(1):6-28.
15. Koopmans L, Bernaards CM, Hildebrandt VH, Van Buuren S, Van der Beek AJ, De Vet HCW. Improving the Individual Work Performance Questionnaire using Rasch Analysis. *Journal of Applied Measurement.* 2014;15(2).
16. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine.* Cambridge University Press; 2011.
17. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. New Jersey: Lawrence Erlbaum Associates; 1988.
18. Schaufeli WB, Bakker AB. *Utrecht Work Engagement Scale: Preliminary Manual.* 2003;Version 1.
19. Schaufeli WB, Bakker AB, Salanova M. The Measurement of Work Engagement With a Short Questionnaire: A Cross-National Study. *Educational and Psychological Measurement.* 2006;66:701-716.
20. Bakker AB, Schaufeli WB, Leiter MP, Taris TW. Work Engagement: An Emerging Concept in Occupational Health Psychology. *Work & Stress.* 2008;22(3):187-200.
21. Demerouti E, Cropanzano R. From Thought to Action: Employee Work Engagement and Job Performance. In: Bakker AB, Leiter MP, eds. *Work Engagement: A Handbook of Essential Theory and Research.* New York: Psychology Press; 2010.
22. Harrison DA, Newman DA, Roth PL. How important are job attitudes? Meta-analytic comparisons of integrative behavioral outcomes and time sequences. *Academy of Management Journal.* 2008;49(2):305-325.
23. Schultz AB, Edington DW. Employee Health and Presenteeism: A Systematic Review. *Journal of Occupational Rehabilitation.* 2007;17:547-579.
24. Locke EA. The nature and causes of job satisfaction. In: Dunnette MD, editor. *Handbook of industrial and organizational psychology.* Palo Alto, CA: Consulting Psychologists Press; 1976:1297-1349.

25. Judge TA, Bono JE, Thoresen CJ, Patton GK. The Job Satisfaction-Job Performance Relationship: A Qualitative and Quantitative Review. *Psychological Bulletin*. 2001;127(3):376-407.
26. Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948.
27. Lerner D, Mosher Henke R. What Does Research Tell Us About Depression, Job Performance, and Work Productivity? *Journal of Occupational and Environmental Medicine*. 2008;50(4):401-410.
28. Kessler RC, Petukhova M, McInnes K, Ustun TB. Content and Scoring Rules for the WHO HPQ absenteeism and presenteeism questions. 2007.
29. Wanous JP, Reichers AE, Hudy MJ. Overall job satisfaction: how good are single-item measures? *J Appl Psychol*. 1997;82(2):247-252.
30. Nagy MS. Using a single-item approach to measure facet job satisfaction. *Journal of Occupational and Organizational Psychology*. 2002;75:77-86.
31. Cunny KA, Perri M. Single-item vs multiple-item measures of health-related quality of life. *Psychological Reports*. 1991;69:127-130.
32. IBM Corp. IBM SPSS Statistics for Windows, Version 20.0. 2011.
33. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*. 2007;60:34-42.
34. Bernaards CM, Proper KI, Hildebrandt VH. Physical Activity, Cardiorespiratory Fitness, and Body Mass Index in Relationship to Work Productivity and Sickness Absence in Computer Workers With Preexisting Neck and Upper Limb Symptoms. *Journal of Occupational and Environmental Medicine*. 2007;49(6):633-640.
35. Fischer CD. Why do lay people believe that satisfaction and performance are correlated? Possible sources of a commonsense theory. *Journal of Organizational Behavior*. 2003;24:753-777.
36. Spector PE, Bauer JA, Fox S. Measurement Artifacts in the Assessment of Counterproductive Work Behavior and Organizational Citizenship Behavior: Do We Know What We Think We Know? *J Appl Psychol*. 2010;97(4):781-790.
37. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology,

- and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*. 2010;63:737-745.
38. Jaramillo F, Carrillat FA, Locander WB. A meta-analytic comparison of managerial ratings and self-evaluations. *Journal of Personal Selling & Sales Management*. 2005;XXV(4):315-328.