Chapter 7

Responsiveness of the Individual Work Performance Questionnaire

Linda Koopmans, Jennifer K. Coffeng, Claire M. Bernaards, Cécile R.L. Boot, Vincent H. Hildebrandt, Henrica C.W. de Vet, Allard J. van der Beek

Article submitted for publication

7

Abstract

Background: Individual work performance is an important outcome measure in studies in the workplace. Nevertheless, its conceptualization and measurement has proven challenging. To overcome limitations of existing scales, the Individual Work Performance Questionnaire (IWPQ) was recently developed. The aim of the current study was to gain insight into the responsiveness of the IWPQ.

Methods: Data were used from the Be Active & Relax randomized controlled trial. The aim of the trial was to investigate the effectiveness of an intervention to stimulate physical activity and relaxation of office workers, on need for recovery. Individual work performance was a secondary outcome measure of the trial. In total, 39 hypotheses were formulated concerning correlations between changes on the IWPQ scales and changes on similar constructs (e.g., presenteeism) and distinct constructs (e.g., need for recovery) used in the trial.

Results: 260 Participants completed the IWPQ at both baseline and 12 months of follow-up. For the IWPQ scales, 23%, 15%, and 38%, respectively, of the hypotheses could be confirmed. In general, the correlations between change scores were weaker than expected. Nevertheless, at least 85% of the correlations were in the expected direction.

Conclusions: Based on results of the current study, no firm conclusions can be drawn about the responsiveness of the IWPQ. Several reasons may account for the weaker than expected correlations. Future research on the IWPQ's responsiveness should be conducted, preferably in other populations and intervention studies, where greater changes over time can be expected.

Introduction

Individual work performance, defined as "*employee behaviors or actions that are relevant to the goals of the organization*" [1], is an important outcome measure in studies in the workplace. Nevertheless, its conceptualization and measurement has proven challenging. First, consensus on a clear definition and conceptual framework of individual work performance (IWP) was lacking. Considering the diversity in conceptual frameworks of IWP, it is not surprising that numerous instruments have been developed to measure (aspects of) IWP. Due to the diversity in conceptual frameworks, the content validity of these existing instruments is questionable (e.g., do they measure the full range of individual work performance?). Also, generic applicability of these instruments is limited, because they are often developed for specific populations (e.g., for a specific occupation, or for workers with health complaints).

To overcome the aforementioned limitations, the Individual Work Performance Questionnaire (IWPQ) was recently developed [2, 3]. The IWPQ is based on a three-dimensional conceptual framework of IWP, which was developed after a systematic review of the literature [4]. This framework includes the dimensions of task performance ("*the proficiency with which individuals perform the core substantive or technical tasks central to his or her job*" [1]), contextual performance ("*behaviors that support the organizational, social and psychological environment in which the technical core must function*" [5]), and counterproductive work behavior ("*behavior that harms the well-being of the organization*" [6]). The IWPQ is a generic instrument, thus, it is suitable for workers in all types of occupations (i.e., blue, pink, and white collar workers) and workers with and without health complaints.

An important purpose of the IWPQ is to assess changes in IWP. For example, we may want to examine fluctuations in IWP over time (e.g., due to age), follow the effects of negative factors on IWP over time (e.g., health problems), or identify successful methods to improve IWP (e.g., intervention studies). In order to do this, the IWPQ must be responsive to changes over time. Responsiveness can be defined as "the ability of an instrument to detect change over time in the construct to be measured" [7]. There is a lot of confusion about the concept over responsiveness, and many different definitions and measures have been proposed over the past decades [8]. In addition, or perhaps, as a result, responsiveness is a seldom examined issue. When assessing responsiveness, we focus on the validity of a *change score*,

which is estimated on the basis of two or more measurement points [8]. The aim of the current study was to gain insight into the responsiveness of the IWPQ.

Methods

Participants

Data were used from the Be Active & Relax "Vitality in Practice" (VIP) randomized controlled trial [9]. The aim of the Be Active & Relax trial was to investigate the effectiveness of an intervention to stimulate physical activity and relaxation of office workers, on need for recovery. By means of stimulating physical activity and relaxation, work-related outcomes (e.g., sickness absenteeism, work engagement and individual work performance) were also expected to improve. The trial included a 2x2 factorial design with four research arms. The four arms consisted of a combined social and physical environmental intervention, a social environmental intervention only, a physical environmental intervention only and a control group. For the purpose of the current study, data of all four groups were taken together. This study was approved by the Medical Ethics Committee of the VU University Medical Center, Amsterdam, The Netherlands. Full details of the design of the Be Active & Relax trial have been reported elsewhere [9].

Measures

Measurements took place at baseline (T0), and at 6 months (T1) and 12 months (T2) follow-up. Only the measurements at baseline and at 12 months (T2) were used to assess responsiveness of the IWPQ.

Individual work performance was measured using the Individual Work Performance Questionnaire (IWPQ) [3, 10]. The IWPQ consists of 18 questions in three scales: task performance, contextual performance, and counterproductive work behavior. The IWPQ 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). The psychometric properties of the IWPQ have been tested and results indicated good to excellent reliability for task performance ($\alpha = 0.78$), contextual performance ($\alpha = 0.85$) and counterproductive work behavior ($\alpha = 0.79$). The IWPQ has shown good face and structural validity [2, 3, 10], as well as sufficient convergent validity and good discriminative validity [11].

Presenteeism was assessed through self-report with the World Health Organization Health and Work Performance Questionnaire (WHO-HPQ) [12].

Presenteeism was assessed by asking participants to rate their actual performance in relation to possible performance. The score represents percentage of performance, and has a lower bound of 0 (*total lack of performance*) and an upper bound of 100 (*top performance*). The reliability and validity of the HPQ was examined for several occupations, and showed good convergent validity. However, poor validity was found for white collar workers [12, 13].

Job satisfaction was assessed using one overall question on a 5-point rating scale from "*highly dissatisfied*" to "*very satisfied*." A single-item measure of job satisfaction has been found to correlate highly with job satisfaction scales, and was therefore considered valid [14, 15].

Work engagement was measured using the Utrecht Work Engagement Scale (UWES) [16]. The UWES consists of three scales (vigour, dedication, and absorption), and a total of 17 items assessed on a 7-polint scale ranging from "*never*" to "*always*." The total score was calculated by adding the means of each scale, and dividing the sum by three. The psychometric properties of this questionnaire have been tested and results indicated an acceptable reliability of vigour (α = 0.68-0.80), dedication (α = 0.91) absorption (α = 0.73-0.75), and the total score (α = 0.93), as well as acceptable convergent validity [16].

Work ability was assessed using one question ("*How do you rate your current work ability compared to lifetime best?*") from the Work Ability Index (WAI) [17], on an 11-point rating scale from 0 "*completely unable to work*" to 10 "*at its best*." The single-item question is very strongly associated with the total WAI, and has shown good predictive validity [18].

Performance rating by the manager was assessed by asking one question ("*How would your manager rate your overall job performance, compared to colleagues in a similar job?")* on a 5-point rating scale from "*much worse*" to "*much better*." This question was adapted from the WHO-HPQ [12] presenteeism question, and previously used in The Netherlands Working Conditions Survey [19]. The reliability and validity of this question is unknown.

Self-rated work quality and quantity were assessed using one question each ("*How do you rate the quality of your own work?" and "How do you rate the quantity of your own work?")* on a 5-point rating scale from "*insufficient*" to "*excellent*." The reliability and validity of these questions is unknown.

Need for recovery (NFR) was assessed using the Need for Recovery after Work scale [20]. This Dutch version of the Questionnaire on the Experience and Evaluation of Work (Dutch abbreviation: VBBA) consists of eleven dichotomous

items (yes/no), representing short-term effects of a day at work. The NFR score is a percentage score (0 to 100) of positive answers of those providing data for at least 8 of the 11 items. The Need for Recovery after Work scale has shown good reliability (α = 0.86-0.88), construct validity, and sensitivity to change in The Netherlands [20-22].

Physical activity was assessed using the Short Questionnaire to Assess Health Enhancing Physical Activity (SQUASH) [23]. Duration and intensity of active commuting, leisure time activities, sport activities, household activities, and physical activities at work (standing and walking), were assessed. For each domain, employees were asked to report the frequency (i.e., times per week), duration of activities (i.e., in minutes), and self-reported intensity (i.e., light, moderate or vigorous). Total scores for minutes per week spent on light, moderate, and vigorous physical activities were calculated. The SQUASH scores have shown reasonable reproducibility (r = 0.57-0.58) and validity against accelerometry (r = 0.45-0.67), which is comparable to other physical activity questionnaires [23, 24].

General health and vitality were measured using the Dutch version of the Rand-36 [25]. General health was measured by asking workers to indicate how they perceived their general health, on a 5-point scale from "*poor*" or "*excellent*." Vitality was measured with a scale of 5 items, asking workers to indicate how often they felt full of life, worn out, tired and full of energy, on a 6-point scale from "*never*" to "*always*." This scale was transformed to a 0-100 score, with higher scores indicating higher vitality. The Dutch version of the Rand-36 has shown good reliability for the vitality scale (α = 0.82) and had reasonable construct validity [25].

Exhaustion was measured using the OLdenburg BurnOut Inventory (OLBI) [26]. The OLBI consists of eight items on a 4-point scale ranging from "*totally disagree*" to "*totally agree*." A mean score was calculated. The OLBI has shown good reliability (α = 0.80-0.85) and reasonable convergent and discriminant validity in different occupational groups [26, 27].

Sickness absenteeism data were retrieved from company records, for the year prior to the intervention (i.e. baseline), and for the year of the intervention (i.e., 12 month follow-up). The score represents the number of workdays absent per year.

Hypotheses

A construct approach of responsiveness testing [8] was applied in the current study, which means that hypotheses were formulated concerning relationships between changes on the IWPQ and changes on other instruments used in the Be Active &

Relax trial. These were divided into hypotheses with similar constructs (e.g., presenteeism) and distinct constructs (e.g., need for recovery). Stronger correlations of the IWPQ scales were hypothesized with constructs similar to IWP than constructs distinct from IWP. Expectations were formulated per IWPQ scale, resulting in a total of 39 hypotheses (3 IWPQ scales x 13 constructs). If positive correlations were expected for task and contextual performance, negative correlations were expected for counterproductive work behavior, and vice versa (also see Table 2).

With similar constructs

The change in each IWPQ scale was expected to correlate moderately (0.30-0.50 or -0.50 - -0.30) with the change in presenteeism [11], job satisfaction [e.g., 28], work engagement [e.g., 29], work ability [e.g., 30], performance rating by the manager [31], work quality, and work quantity. Based on literature, the change in counterproductive work behavior was expected to correlate weakly or not at all (-0.20 – 0.20) with the change in the last three constructs [32].

With distinct constructs

The change in each IWPQ scale was expected to correlate weakly (0.20 – 0.30 or -0.30 –-0.20) with the change in need for recovery [e.g., 27, 33], physical activity [e.g., 34], general health [e.g., 35, 36], vitality [e.g., 37], and exhaustion [e.g., 38]. Finally, the change in each IWPQ scale was expected to correlate weakly or not at all (-0.20 – 0.20) with the change in sickness absenteeism [39, 40].

Data analysis

Pearson correlations between the change scores of each IWPQ scale and the change scores on the other constructs were calculated for the change scores from baseline (T0) to 12 months (T2). Only participants who completed the IWPQ at both T0 and T2 were included in the data analysis. Analyses were conducted in SPSS 20.0 [41].

Results

Descriptive statistics of the participants

Of the 412 participants in the Be Active & Relax trial, 260 participants (63%) completed the IWPQ at both baseline and 12 months. At baseline (n=260), participants had a mean age of 43.2 years (SD = 9.9), and 37% was female.

Descriptive statistics of the IWPQ scales and the other constructs

Table 1 presents the mean scores and standard deviations (SD) on the IWPQ scales and the other constructs at baseline (T0) and 12 months (T2). It also reports the mean and standard deviation (SD_{change}) of the change scores on the IWPQ scales and the other constructs from T0 to T2.

Table 1. Mean scores (and SD) and mean change scores (and SD_{change}) on the IWPQ scales and the similar/distinct constructs at baseline (T0) and 12 months (T2)

	T0 (baseline) Mean (SD)	**T2 (12 months)** Mean (SD)	**Change score T2-T0** Mean (SD_{change})
IWPQ (1-5)			
Task performance	3.46 (0.68)	3.63 (0.66)	0.17 (0.70)
Contextual performance	3.34 (0.71)	3.39 (0.79)	0.04 (0.69)
Counterproductive work behavior	2.23 (0.65)	2.16 (0.66)	-0.07 (0.64)
Similar constructs			
Presenteeism (0-100)	76.58 (8.76)	75.87 (10.62)	-0.79 (11.51)
Job satisfaction (1-5)	3.96 (0.73)	3.85 (0.75)	-0.11 (0.80)
Work engagement (1-7)	4.91 (0.85)	4.84 (0.93)	-0.07 (0.71)
Work ability (1-10)	7.79 (1.42)	7.70 (1.57)	-0.08 (1.56)
Performance rating by the manager (1-5)	3.41 (0.81)	3.46 (0.81)	0.06 (0.81)
Self-rated work quality (1-5)	3.83 (0.79)	3.63 (0.87)	-0.20 (0.95)
Self-rated work quantity (1-5)	3.87 (0.83)	3.74 (0.92)	-0.12 (0.95)
Distinct constructs			
Need for recovery (0-100)	32.20 (29.26)	27.78 (28.71)	-2.40 (23.70)
Physical activity (min/week)			
Light	1810.10 (1363.68)	1603.23 (1618.94)	-199.40 (1785.64)
Moderate	281.81 (254.19)	350.94 (633.98)	72.66 (629.00)
Vigorous	83.53 (160.15)	99.79 (272.90)	9.40 (266.15)
General health (1-5)	3.35 (0.85)	3.37 (0.84)	0.79 (1.53)
Vitality (0-100)	64.08 (18.84)	65.72 (17.97)	1.87 (15.17)
Exhaustion (1-4)	2.15 (0.48)	2.15 (0.46)	0.04 (0.40)
Sickness absenteeism (workdays absent per year)	7.55 (21.81)	7.37 (20.91)	0.55 (25.03)

Correlations between change scores

Table 2 presents the expected and observed correlations between the change scores of the IWPQ scales and the change scores of the other constructs. For task performance, 85% of the correlations were in the expected direction, and for contextual performance and counterproductive work behavior, 92% of the correlations were in the expected direction. However, in many cases, the correlations were weaker than expected.

For the task performance scale, 3 out of 13 (23%) hypotheses were fully confirmed. As expected, the change in task performance correlated moderately positive with the changes in vitality (r = 0.23), moderately negatively with the change in exhaustion (r = -0.23), and weakly negative with the change in absenteeism (r = -0.14).

For the contextual performance scale, 2 out of 13 (15%) hypotheses were fully confirmed. As expected, the change in contextual performance correlated moderately positive with the change in vitality (r = 0.29), and weakly negative with the change in absenteeism (r = -0.08). Furthermore, the correlation between the change in contextual performance and the changes in most of the similar constructs (e.g., presenteeism, work engagement, work ability) approached the 0.30 correlation strength.

For the counterproductive work behavior scale, 5 out of 13 (38%) hypotheses were fully confirmed. As expected, the change in counterproductive work behavior correlated weakly with the changes in rating by the manager (r = -0.02), work quality (r = -0.06), work quantity (r = 0.02), and absenteeism (r = -0.09), and moderately positive with the change in exhaustion (r = 0.23).

Table 2. Pearson correlations (E = expected, O = observed) between **change scores** of the IWPQ scales and similar/distinct constructs

	IWPQ scale		
	Task performance	**Contextual performance**	**Counterproductive work behavior**
Similar constructs			
Presenteeism	E: 0.30 – 0.50 O: 0.18	E: 0.30 – 0.50 O: 0.22	E: -0.50 – -0.30 O: -0.11
Job satisfaction	E: 0.30 – 0.50 O: 0.12	E: 0.30 – 0.50 O: 0.17	E: -0.50 – -0.30 O: -0.24
Work engagement	E: 0.30 – 0.50 O: 0.19	E: 0.30 – 0.50 O: 0.29	E: -0.50 – -0.30 O: -0.23
Work ability	E: 0.30 – 0.50 O: 0.16	E: 0.30 – 0.50 O: 0.26	E: -0.50 – -0.30 O: -0.23
Performance rating by the manager	E: 0.30 - 0.50 O: 0.16	E: 0.30 - 0.50 O: 0.22	E: -0.20 – 0.20 O: -0.02 *
Work quality	E: 0.30 – 0.50 O: 0.20	E: 0.30 – 0.50 O: 0.18	E: -0.20 – 0.20 O: -0.06 *
Work quantity	E: 0.30 – 0.50 O: 0.11	E: 0.30 – 0.50 O: 0.19	E: -0.20 – 0.20 O: 0.02 *
Distinct constructs			
Need for recovery	E: -0.30 – -0.20 O: -0.15	E: -0.30 – -0.20 O: -0.11	E: 0.20 – 0.30 O: 0.16
Physical activity	E: 0.20 – 0.30	E: 0.20 – 0.30	E: -0.30 – -0.20
Light	O: -0.09	O: -0.04	O: -0.07
Moderate	O: 0.03	O: 0.03	O: -0.07
Vigorous	O: -0.05	O: 0.00	O: -0.04
General health	E: 0.20 – 0.30 O: -0.07	E: 0.20 – 0.30 O: 0.08	E: -0.30 – -0.20 O: 0.02
Vitality	E: 0.20 – 0.30 O: 0.23 *	E: 0.20 – 0.30 O: 0.29 *	E: -0.30 – -0.20 O: -0.03
Exhaustion	E: -0.30 – -0.20 O: -0.23 *	E: -0.30 – -0.20 O: -0.13	E: 0.20 – 0.30 O: 0.23 *
Sickness absenteeism	E: -0.20 - 0.20 O: -0.14 *	E: -0.20 - 0.20 O: -0.08 *	E: -0.20 - 0.20 O: -0.09 *
Hypotheses:			
Confirmed	23%	15%	38%
In the right direction	85%	92%	92%

Note: E = expected correlation, O = observed correlation. * = Confirmed hypothesis.

Discussion

The aim of the current study was to examine the responsiveness of the IWPQ, i.e., the ability of the IWPQ to detect change over time. A total of 39 hypotheses were formulated concerning the relationships between changes on the IWPQ and changes on similar constructs (e.g., presenteeism) and distinct constructs (e.g., need for recovery) used in the Be Active & Relax trial. For the IWPQ task performance, contextual performance, and counterproductive work behavior scales, 23%, 15%, and 38%, respectively, of the hypotheses could be confirmed. As hypothesized, the correlations of the IWPQ scales were slightly stronger with similar constructs than with distinct constructs, on average. However, in general, the correlations between change scores were weaker than expected. Nevertheless, most of the correlations (at least 85%) were in the expected direction. Exceptions were the correlations between the change scores of task performance and light and intense physical activity ($r = -0.09$ and -0.05, respectively), task performance and general health ($r = -0.07$), contextual performance and light physical activity ($r = -0.04$), and counterproductive work behavior and general health ($r = 0.02$).

Several reasons may account for the weaker than expected correlations. First, the IWPQ questions may not be sensitive enough to pick up changes in IWP over time. Also, it is hard to say how a change from answer categories “regularly” to “often” can be achieved. What needs to be done to accomplish a change from “regularly” to “often,” e.g., in keeping your work results in mind? And what does this change mean? In sum, the questions of the IWPQ scales may lack discriminative ability. However, in the developmental phase of the IWPQ scales, Rasch analysis [42] was performed to make sure that those items with a high discrimination parameter (i.e., high slope) were retained in the IWPQ 1.0 [2, 3]. Also, in the construct validation phase of the IWPQ scales, the IWPQ 1.0 was able to differentiate between known groups [11]. This suggests that the items in the IWPQ scales should have enough discriminative ability to detect changes in IWP over time.

Possibly, low responsiveness of the IWPQ could be caused by ceiling and floor effects in the scales. Although previous examination of the IWPQ using Rasch analysis has shown that the items of the IWPQ are relatively well-distributed over the scales, persons continue to score relatively high on task performance (ceiling effect), and low on CWB (floor effect) [3]. This could be caused by the tendency of persons to evaluate and present themselves in a socially desirable, favorable way [43, 44]. As a consequence of the ceiling and floor effects, it becomes hard to detect

further improvements in task performance, and further decreases in CWB. Thus, the ability to detect changes at the high part of the task performance scale, and low part of the CWB scale, may be diminished.

Another possible reason for the lower than expected correlations may lie in the study population. As said before, the population in the current study consisted of relatively healthy, well-functioning office workers who, in general, scored high on constructs such as general health, presenteeism, and job satisfaction, and low on constructs such as need for recovery, exhaustion, and sickness absenteeism. This makes it hard to obtain or detect any further improvements in this population. Despite the use of an intervention, small changes on the constructs over the 12-month intervention period were obtained. When examining the scatterplots of the change scores, low spread on many constructs can be observed (i.e., dots clustered in the middle), and this can cause deflated correlations [8].

Finally, a reason for the lower than expected correlations may be that the intervention was not effective enough to obtain changes in IWP. The primary aim of the Be Active & Relax study was to investigate the effectiveness of an intervention to stimulate physical activity and relaxation of office workers, on need for recovery [9]. Indirectly, physical activity and relaxation were expected to improving IWP. However, and it may be that the intervention was not specific or intense enough to obtain improvements in IWP. Despite the fact that the intervention was not directly targeted at IWP, and despite high baseline levels on the constructs, a statistically significant increase in tasks performance (B = 0.2, 95% CI 0.0; 0.4), and a statistically significant decrease in contextual performance (B = -0.3, 95% CI -0.4; 0.1), were detected in the Be Active & Relax study [45]. The decrease in contextual performance could be explained by the fact that participants in the intervention groups were stimulated to engage in physical activity and relaxation during the workday, and this possibly could have reduced taking on extra work tasks, for example. Thus, this study showed that the IWPQ is able to detect statistically significant changes in individual work performance over time.

Assessment of responsiveness

As stated in the Introduction, there is a lot of confusion about the concept over responsiveness, and many different definitions and measures have been proposed over the past decades [8]. In addition, or perhaps, as a result, responsiveness is a seldom examined issue. For example, Abma et al. [46] reviewed the measurement properties of five self-report (health-related) work functioning instruments; the

EWPS, WLQ, SPS, WPS, and LEAPS. For all five instruments, the methodological quality of responsiveness testing was poor, or not studied. Of the instruments used in the current study, only the responsiveness of the Need for Recovery Scale was examined. Based on effect sizes, the responsiveness of this scale appeared to be good [21]. However, the responsiveness of the other questionnaires used in the current study remains unknown. This is a limitation of the responsiveness testing process, because responsiveness of a new questionnaire is tested against change scores of existing questionnaires, whose responsiveness is also unknown, and may be poor.

No golden standard or clear guidelines seem to exist for the assessment of responsiveness and the interpretation of results. De Vet and colleagues [8] stated that responsiveness is often examined based on inappropriate outcome measures, such as effect sizes or standardized response mean. They advise that responsiveness should be seen as a form of longitudinal validity, using either a criterion approach (if a gold standard is available) or a construct approach (testing hypotheses of change scores).

In addition to the lack of clarity on how responsiveness should be tested, there are no clear guidelines as to what the strength of correlations between change scores should be. A final reason for the large percentage of unconfirmed hypotheses in the current study, may be that the hypothesized correlations (r = 0.30-0.50) were too high to begin with. In line with Cohen [47], we interpreted a correlation coefficient over 0.50 as strong, 0.30 to 0.50 as moderate, 0.10 to 0.30 as weak, and below 0.10 as no relation between constructs at all. Often, Cohen's guidelines are used for cross-sectional correlations, i.e., when a correlation between two different measurement scores obtained at the same point in time is examined (thus, there is only one measurement). When it comes to correlations between change scores (multiple measurements), it is based on two measurements, and a double measurement error is involved. Due to this double measurement error, it seems reasonable that lower correlations may be expected. This issue has been addressed by other researchers. For example, Abma et al. [48] examined the responsiveness of the Work Role Functioning questionnaire, and they hypothesized correlation sizes around 0.20 to 0.30 with other constructs, because it was expected that many participants would show no changes, and based on results in earlier studies with similar questionnaires. For the constructs used in the current study, previous research has shown that, for example, the cross-sectional correlation between IWP and work engagement ranges between r = 0.30-0.50 [e.g., 49]. It is therefore

questionable whether correlations of r = 0.30-0.50 between their change scores can reasonably be expected. Such high correlations between change scores would likely be obtained for identical constructs, rather than similar (but not identical) constructs.

Recommendations for future research

The responsiveness of the IWPQ should be further examined in future research, to determine whether its responsiveness is truly low, or whether the low responsiveness found in the current study was caused by limitations of the current study. We therefore recommend examining the responsiveness of the IWPQ in different populations, preferably in populations with low(er) baseline levels on the constructs, where large(r) changes on the constructs over time can be expected. Suggestions for such populations could be a sample of workers with work-related musculoskeletal health problems, mental health problems, and/or low job satisfaction. An intervention study, which is directly aimed at improving IWP, could obtain greater changes in these populations, making it easier to detect changes in IWP and related constructs. Suggestions for such a study could be an intervention focusing on managerial style, technological improvements at work, and/or job skills training. Also, the responsiveness of the IWPQ should preferably be examined using other measurement instruments of which the responsiveness is known. Finally, the responsiveness of questionnaires deserves greater attention, and clear guidelines for assessing and interpreting responsiveness should be adopted. The guidelines proposed by Terwee et al. [50], Mokkink et al. [51], and De Vet et al. [8] could provide a good starting point for this.

Conclusion

Based on results of the current study, no firm conclusions can be drawn about the responsiveness of the IWPQ. Overall, most of the correlations between changes on the IWPQ scales and changes on other constructs were in the expected direction, although not as high as expected. This might indicate low responsiveness of the IWPQ. However, the weaker than expected correlations may also be accounted for by characteristics of the intervention study, such as the relatively healthy, well-functioning study population, and an intervention study that was not primarily aimed at IWP. Nevertheless, the IWPQ was able to show statistically significant changes in IWP during baseline and 12 months follow-up. Future research should provide more information about the responsiveness of the IWPQ, preferably in other populations and intervention studies.

References

1. Campbell JP. Modeling the performance prediction problem in industrial and organizational psychology. In Handbook of industrial and organizational psychology. Volume 1. 2nd edition. Edited by Dunnette MD & Hough, LM. (1990). Palo Alto, CA, US: Consulting Psychologists Press; 1990:687-755.
2. 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.
3. 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).
4. Koopmans L, Bernaards CM, 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):856-66.
5. Borman WC, Motowidlo SJ. Expanding the criterion domain to include elements of contextual performance. In: Personnel Selection in Organizations. Edited by Schmitt N, Borman WC. San Francisco, CA: Jossey Bass; 1993. p. 71-98.
6. 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.
7. 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;63:737-45.
8. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine. Cambridge University Press; 2011.
9. Coffeng JK, Hendriksen IJM, Duijts SF, Proper KI, Van Mechelen W, Boot CRL. The development of the be active & relax vitality in practice (VIP) project and design of an RCT to reduce the need for recovery in office employees. BMC Public Health 2012;12:592.

10. 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).
11. Koopmans L, Bernaards CM, Hildebrandt VH, De Vet HCW, Van der Beek AJ. Construct validity of the individual work performance questionnaire. Journal of Occupational and Environmental Medicine, 2014;56(3).
12. Kessler RC, Barber C, Beck A, Berglund P, Cleary PD, McKenas D, Pronk N, Simon G, Ustun TB, Wang P. The world health organization health and work performance questionnaire (HPQ). Journal of Occupational and Environmental Medicine 2003;45:156-74.
13. Kessler RC, Ames M, Hymel PA, Loeppke R, McKenas DK, Richling DE, Stang PE, Ustun TB. Using the world health organization health and work performance questionnaire (HPQ) to evaluate the indirect workplace costs of illness. Journal of Occupational and Environmental Medicine 2004;46(6):S23-37.
14. Wanous J.P., Reichers AE, Hudy MJ. Overall job satisfaction: How good are single-item measures? Journal of Applied Psychology 1997;82(2):247-52.
15. Nagy MS. Using a single-item approach to measure facet job satisfaction. Journal of Occupational and Organizational Psychology 2002;75:77-86.
16. Schaufeli WB, Bakker AB. Utrecht Work Engagement Scale: Preliminary Manual. Occupational Health Psychology Unit, Utrecht University; 2003. Version 1.
17. Ilmarinen J. The work ability index (WAI). Occupational medicine (Oxford, England). 2007;57:160.
18. Ahlstrom L, Grimby-Ekman A, Hagberg M, Dellve L. The work ability index and single-item question: Associations with sick leave, symptoms, and health - a prospective study of women on long-term sick leave. Scandinavian Journal of Work, Environment and Health 2010;36(5):404-12.
19. Koppes LLJ, De Vroome EMM, Mol MEM, Janssen BJM, Van den Bossche SNJ. The Netherlands working conditions survey. PlantijnCasparie, Almere: The Netherlands Organisation for Applied Scientific Research; 2008.
20. Van Veldhoven M, Broersen S. Measurement quality and validity of the "need for recovery scale". Occupational and environmental medicine 2003;60(Suppl 1):i3-9.

21. De Croon EM, Sluiter JK, Frings-Dresen MH. Psychometric properties of the need for recovery after work scale: Test-retest reliability and sensitivity to detect change. Occupational and environmental medicine 2006;63(3):202-6.
22. Van Veldhoven MJ, Sluiter JK. Work-related recovery opportunities: Testing scale properties and validity in relation to health. International Archive of Occupational and Environmental Health 2009;82(9):1065-75.
23. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. Journal of clinical epidemiology 2003;56(12):1163-9.
24. Wagenmakers R, van den Akker-Scheek I, Groothoff JW, Zijlstra W, Bulstra SK, Kootstra JW, Wendel-Vos GC, Van Raaij JJ, Stevens M. Reliability and validity of the short questionnaire to assess health-enhancing physical activity (SQUASH) in patients after total hip arthroplasty. BMC Musculoskelet.Disord. 2008;17(9):141.
25. Van der Zee KI, Sanderman R. Het meten van de algemene gezondheidstoestand met de rand-36: Een handleiding (2nd ed.). Groningen: Noordelijk Centrum voor Gezondheidsvraagstukken, NCG; 2012.
26. Demerouti E, Bakker AB, Vardakou I, Kantas A. The convergent validity of two burnout instruments: A multitrait-multimethod analysis. European Journal of Psychological Assessment 2003;19(1):12-23.
27. Demerouti E, Bakker AB. The oldenburg burnout inventory: A good alternative to measure burnout and engagement. In: Halbesleben JRB, editor. Handbook of Stress and Burnout in Health Care. Happauge, NY: Nova Science; 2008.
28. 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.
29. Bakker AB, Demerouti E. Towards a model of work engagement. Career Development International 2008;13(3):209-23.
30. Van den Berg TIJ. The role of work ability and health on sustaining employability [dissertation]. Erasmus University Rotterdam; 2010.
31. Harris MM, Schaubroeck J. A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. Personnel Psychology 1988;41:43-62.
32. Dalal RS. A meta-analytic comparison of managerial ratings and self-evaluations. Journal of Personal Selling & Sales Management 2005;XXV(4):315-28.

33. Fritz C, Sonnentag S. Recovery, health, and job performance: Effects of weekend experiences. Journal of occupational health psychology 2005;10(3):187-99.
34. Pronk NP, Martinson B, Kessler RC, Beck AL, Simon GE, Wang P. The association between work performance and physical activity, cardiorespiratory fitness, and obesity. Journal of Occupational and Environmental Medicine 2004;46(1):19-25.
35. Schultz AB, Edington DW. Employee health and presenteeism: A systematic review. Journal of Occupational Rehabilitation 2007;17:547-79.
36. Boles M, Pelletier B, Lynch W. The relationship between health risks and work productivity. Journal of Occupational and Environmental Medicine 2004;46(7):737-45.
37. Wright TA, Bonett DG, Sweeney DA. Mental health and work performance: Results of a longitudinal field study. Journal of Occupational and Organizational Psychology 1993;66:277-84.
38. Wright TA, Cropanzano R. Emotional exhaustion as a predictor of job performance and voluntary turnover. Journal of Applied Psychology 1998;83(3):486-93.
39. Bycio P. Job performance and absenteeism: A review and meta-analysis. Human Relations 1992;45(2):193-220.
40. Morrow PC, McElroy JC, Laczniak KS, Fenton JB. Using absenteeism and performance to predict employee turnover: Early detection through company records. Journal of vocational behavior 1999;55:358-74.
41. IBM Corp. IBM SPSS statistics for windows, version 20.0. 2011.
42. Rasch G. Probabalistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1960.
43. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology 2003;88(5):879-903.
44. Van der Heijden BIJM, Nijhof AHJ. The value of subjectivity: Problems and prospects for 36-degree appraisal systems. The International Journal of Human Resource Management 2004;15(3):493-511.
45. Coffeng JK, Hendriksen IJM, Duijts SF, Proper KI, Van Mechelen W, Boot CRL. Effectiveness of a combined social and physical environmental intervention on work-related outcomes in office employees, submitted for publication.

46. Abma FI, Van der Klink JJL, Terwee CB, Amick III BC, Bultmann U. Evaluation of the measurement properties of self-reported health-related work-functioning instruments among workers with common mental disorders. Scandinavian Journal of Work, Environment and Health 2012;38(1):5-18.
47. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New Jersey: Lawrence Erlbaum Associates; 1988.
48. Abma FI, van der Klink JJ, Bültmann U. The work role functioning questionnaire 2.0 (Dutch version): Examination of its reliability, validity and responsiveness in the general working population. Journal of Occupational Rehabilitation 2013;23(1):135-47.
49. Demerouti E, Bakker AB. Employee well-being and job performance: Where we stand and where we should go. In: Houdmont J, McIntyre S, editors. Occupational Health Psychology: European Perspectives on Research, Education and Practice. 1st ed. Maia: ISMAI Publications; 2006.
50. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. Quality of Life Research 2003;12(4):349-62.
51. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, Knol DL, Bouter LM, De Vet HCW. Inter-rater agreement and reliability of the COSMIN (COnsensus-based standards for the selection of health status measurement instruments) checklist. BMC Medical Research Methodology 2010;10(82).