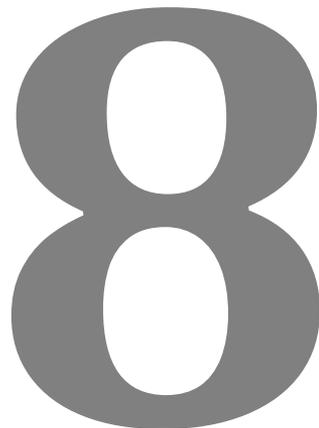


**Cross-Cultural Adaptation of the Individual  
Work Performance Questionnaire**

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

Article submitted for publication



## Abstract

**Objectives:** The study objectives were to perform a cross-cultural adaptation of the Individual Work Performance Questionnaire (IWPQ) from the Dutch to the American-English language, and to assess the questionnaire's internal consistency and content validity in the American-English context.

**Methods:** The Dutch IWPQ was translated and adapted to the American-English language using guidelines of Beaton et al. (2000). The process consisted of five steps: a forward translation by two independent translators, synthesis, back-translation by two other independent translators, an expert committee review, and pilot-testing. During the pilot-testing, cognitive interviews with 40 American workers were performed, to examine the comprehensibility, applicability, and completeness of the American-English IWPQ.

**Results:** The questionnaire translation was conducted without major difficulties. The questionnaire instructions were slightly modified to aid interpretation in the American-English language. Inconsistencies with verb tense were identified, and it was decided to consistently use the simple past for the American-English version. In general, participants were positive on the comprehensibility, applicability and completeness of the questionnaire during the pilot-testing phase. Items TP3 and TP4 might warrant further attention. Furthermore, the study showed good results concerning the internal consistency of the American-English IWPQ (Cronbach's alphas for the scales between 0.79 and 0.89) and good content validity.

**Conclusion:** The results indicate that the cross-cultural adaptation of the American-English IWPQ was successful and that the measurement properties of the translated version are good.

## Introduction

In today's world, it is increasingly important to maintain, improve, and optimize individual work performance (IWP) of employees. In Europe and the United States of America, for example, the 'grey wave' (i.e., accelerated growth of the older working population and a decline in numbers of the younger working population) and the economic recession force companies and employees to perform more or better work with less people. Also, due to the grey wave, the retirement age of older workers has been prolonged [e.g., 1]. Thus, their IWP has to be maintained until a later age. In order to accurately establish the effectiveness of interventions, procedures and strategies to maintain, improve, or optimize IWP, valid measurement of IWP is a prerequisite.

IWP, defined as "*behaviors or actions that are relevant to the goals of the organization*", is since long considered to be a multidimensional construct [2,3]. Based on several reviews of the literature [4-6], it can be concluded that IWP consists of three broad dimensions: task performance, contextual performance, and counterproductive work behavior. The first dimension, *task performance*, traditionally has received most attention, and can be defined as "the proficiency with which individuals perform the core substantive or technical tasks central to their job" [2]. The second dimension of IWP is *contextual performance*, defined as "behaviors that support the organizational, social and psychological environment in which the technical core must function" [7]. The third dimension of IWP is *counterproductive work behavior*, defined as "behavior that harms the well-being of the organization" [5].

Recently, the Individual Work Performance Questionnaire (IWPQ) [8,9] was developed in The Netherlands. The IWPQ is the first questionnaire to incorporate all relevant dimensions of IWP into one questionnaire. An advantage of this is that the content of each scale is fitted to the content of the other scales. As a result, the scales do not include antithetical items, that is, items overlapping in content [10]. Another advantage of the IWPQ is that it is generically applicable. Previous questionnaires to measure IWP were often developed for, or refined based on, specific populations, such as workers in specific jobs [e.g., 11,12] or workers with health problems [e.g., 13,14]. The IWPQ can be used for workers in blue, pink, and white collar jobs, and workers with and without health problems [e.g., 8,9].

Considering the advantages of the IWPQ as a comprehensive and generic measurement instrument of IWP, it seems especially suitable for examining the

effectiveness of interventions, procedures and strategies to maintain, improve, or optimize IWP. In order for the IWPQ to be used outside of The Netherlands, it has to be cross-culturally adapted and validated. Because of possible cultural differences between countries, instruments need to be systematically translated, adapted, and validated before they can be used in other cultural contexts. Beaton et al. [15] have proposed a guideline for cross-cultural translation and adaptation, that consists of five steps: a forward translation by two independent translators, synthesis, back-translation by two other independent translators, an expert committee review, and pilot-testing. In the pilot-testing phase, cognitive interviews are held with people from the target population, in order to get an understanding of the comprehensibility, applicability, and completeness of the translated questionnaire. The objectives of the current study were to perform a cross-cultural adaptation of the Individual Work Performance Questionnaire (IWPQ) from the Dutch to the American-English language, and to assess the questionnaire's internal consistency and content validity in the American-English context.

## Methods

### Individual Work Performance Questionnaire

The Individual Work Performance Questionnaire (IWPQ) [8,9] measures *“employee behaviors or actions that are relevant to the goals of the organization”* [2]. The IWPQ consists of 18 items, divided into three scales: task performance, contextual performance, and counterproductive work behavior (see Table 1). All items have 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). A mean score for each IWPQ scale can be calculated by adding the item scores, and dividing their sum by the number of items in the scale. Hence, the IWPQ yields three scale scores that range between 0 and 4, with higher scores reflecting higher task and contextual performance, and higher counterproductive work behavior.

### Cross-cultural adaptation

The IWPQ's cross-cultural adaptation process followed the guidelines of Beaton et al. [15], pictured in Figure 1.

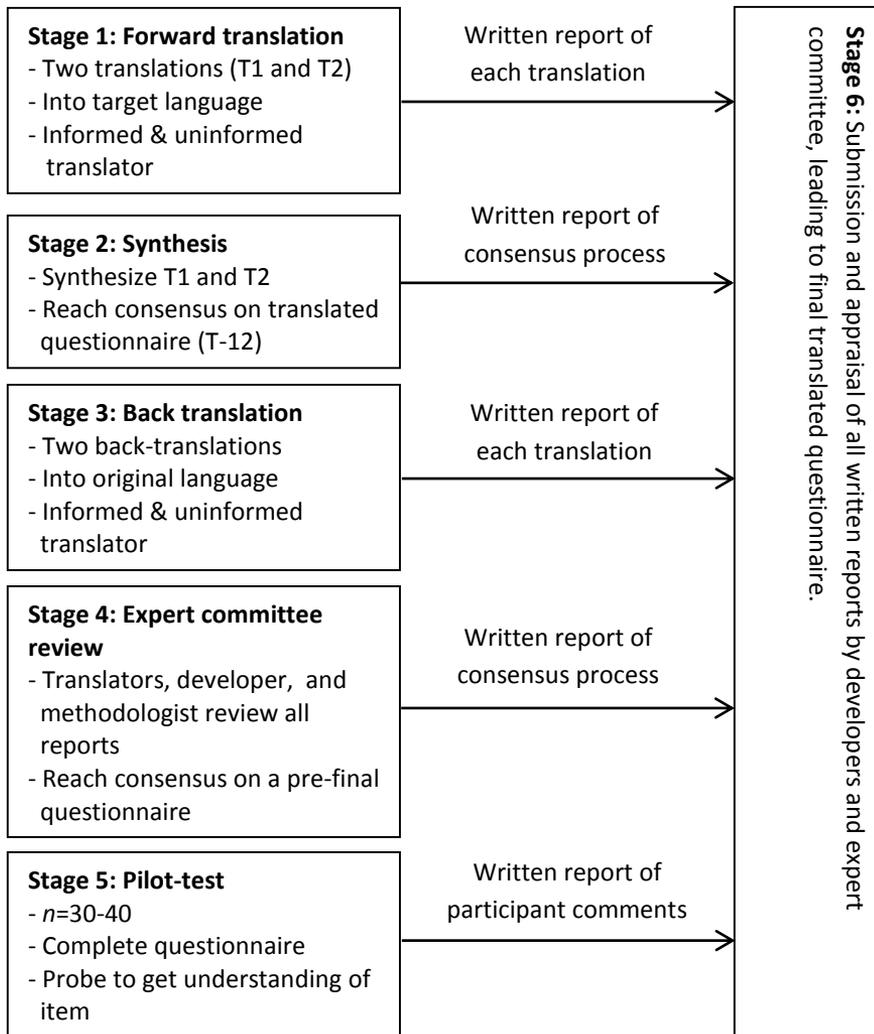


Figure 1. Recommended stages of the cross-cultural adaptation process, based on Beaton et al. [15].

### *Forward translation*

The forward translation of the IWPQ's instruction, items, and answer categories, was performed by two independent translators. Both translators were bilingual, with American-English as their mother tongue. One translator had expertise on individual work performance, and the other translator was naive about the topic. Both translators wrote a report of the translation, containing challenging phrases and uncertainties, and considerations for their decisions.

### *Synthesis*

The results of both translations (T1 and T2) were compared by the two translators and one researcher (LK). A written report documented the consensus process, the discrepancies, and how the discrepancies were resolved. The translators and the researcher reached consensus on one common American-English questionnaire (T-12).

### *Back translation*

The common American-English questionnaire was back-translated into Dutch by two other independent translators. Both translators were bilingual, with Dutch as their mother tongue. One translator was naive about the topic, whereas the other translator had expertise on the topic. Both translators wrote a report of the translation, containing challenging phrases and uncertainties, and considerations for their decisions.

### *Expert committee review*

All the translated versions were combined into one pre-final questionnaire by an expert committee. The expert committee consisted of the four translators, one researcher (LK), and one methodologist (HdV). Discrepancies between the original and translated versions were identified and discussed. Also, semantic, idiomatic, experiential and conceptual equivalences were evaluated. Again, a written report documented the consensus process, the discrepancies, and how the discrepancies were resolved. The expert committee reached consensus on a pre-final American-English version of the IWPQ.

### *Pilot-test*

To examine the comprehensibility, applicability, and completeness of the translated questionnaire, a pilot-test was performed. A total of 40 participants were included

in the pilot-test. Inclusion criteria were: currently working (8 hours a week or more), aged 18-65 years, and able to read and understand the American-English language. Participants were recruited among employees of Tufts Medical Center in Boston, MA. In order to promote participation in the pilot-test, an outreach e-mail was sent to employees of participating departments, after which an appointment with the researcher (LK) could be made. The pilot-test was approved by the Institutional Review Board of Tufts Medical Center (IRB number 10929).

After signing an informed consent file, participants filled in the American-English IWPQ. “Think aloud” and “probing” techniques [16] were used in order to identify participants’ opinion on the comprehensibility, applicability, and completeness of the instructions, items, and answer categories of the translated questionnaire. The duration of the pilot-test was on average 15 minutes, including questionnaire completion. Participants’ comments were written down into a report by the researcher (LK). The comments were independently assessed by two researchers (LK and CB), after which a consensus meeting took place. Any discrepancies that remained were discussed with the translators and the other IWPQ developers (VH, HdV, and AvdB), after which consensus was reached on a final American-English questionnaire.

### **Measurement properties of the pre-final questionnaire**

Descriptive statistics of the IWPQ items and scales, and of the socio-demographic characteristics of the participants (gender, age, number of work hours a week, and primary type of occupation) were used to examine the distribution of the data. Internal consistency of the IWPQ scales was determined using Cronbach’s alpha [17]. Item-to-scale correlations were calculated to evaluate the fit of the item within the scale. Furthermore, scale scores were examined for floor or ceiling effects (> 15% at the extreme values [16]). Statistical analyses of the data were done in SPSS20.

The content validity of the American-English questionnaire was evaluated by the members of the expert committee throughout the cross-cultural adaptation process, and by the developers of the IWPQ through qualitative analysis of the comments provided by the participants of the pilot-test.

## Results

### Cross-cultural adaptation

#### *Translation*

The forward translation of the IWPQ was conducted and some challenging issues were encountered. All issues were discussed among the two translators and the researcher, until consensus emerged. First, conceptual issues were identified with the instruction. "Behavior at work" was considered too evaluative, and might imply whether or not you obeyed rules like a child. To obtain conceptual equivalence to the original meaning, it was chosen to use "how you conducted yourself at work." Second, for some questionnaire items, inconsistencies with the verb tense were identified. In Dutch, the simple past (e.g., "started") and the present perfect (e.g., "have started") are used interchangeably. It was chosen to consistently use the simple past in the American-English version, because the items refer to a completed action in the past 3 months. Furthermore, there were some idiomatic issues in the translation of items TP1, 3 and 4 ("I was able to" versus "I succeeded in"), TP2 ("keep in mind" versus "keep in sight"), CP8 and 9 ("keep up-to-date" versus "maintain"), CWB14 ("issues" versus "things") and CWB16-18 ("aspects" versus "sides"). The main reasons for choosing the first option for each item, were similarity to the original Dutch item, generic applicability (suitable for workers in all types of jobs), and appropriateness of a word (decent, proper).

The back-translation was conducted without major difficulties. Issues were discussed among the members of the expert committee, until consensus emerged. First, a conceptual issue was identified with the instruction sentence "how you conducted yourself at work." Comment was that you cannot "conduct yourself." To obtain conceptual equivalence to the original meaning, it was chosen to use "how you carried out your work." Second, there were some linguistic and conceptual issues in the wording of items TP2 ("results I needed to achieve in my work" was considered incorrect use of American-English), TP3 ("distinguish between" was considered double use of words), TP4 ("perform my work" versus "conduct my work"), and CP6 ("I started new tasks on my own initiative" was considered double use of words). These four items were adapted to correct use of American-English, and to obtain conceptual equivalence to the original meaning. Lastly, one translator expressed issues with the answer category "seldom," and wondered whether this should be "seldomly" or "rarely." However, consensus emerged that "seldom" is often used in answer scales of American-English questionnaires, and is correct use of American-English. Table 1 shows the pre-final American-English IWPQ.

Table 1. The pre-final American-English IWPQ and descriptive statistics of the items

Instructions: The following questions relate to how you carried out your work during the past 3 months. In order to get an accurate picture of your conduct at work, it is important that you complete the questionnaire as carefully and honestly as possible. If you are uncertain about how to answer a particular question, please give the best possible answer. The questionnaire will take about 5 minutes to complete. The questionnaire is completely anonymous: your answers will not be seen by your supervisor(s) or colleagues.

Task performance (TP) scale		Response, n (%)				Mean (SD)	Item-to-scale correlation	
		0 (seldom)	1 (sometimes)	2 (regularly)	3 (often)			4 (always)
In the past 3 months...								
TP1	I was able to plan my work so that I finished it on time. *	0 (0)	7 (17.5)	9 (22.5)	13 (32.5)	11 (27.5)	2.70 (1.07)	0.64
TP2	I kept in mind the work result I needed to achieve. *	0 (0)	1 (2.5)	7 (17.5)	12 (30.0)	20 (50.0)	3.28 (0.84)	0.58

Table 1. Continued

Task performance (TP) scale		Response, n (%)				Mean (SD)	Item-to-scale correlation	
		0 (seldom)	1 (sometimes)	2 (regularly)	3 (often)			4 (always)
In the past 3 months...								
TP3	I was able to distinguish main issues from side issues. * §	1 (2.5)	10 (25.0)	15 (37.5)	10 (25.0)	4 (10.0)	3.08 (0.76)	0.46
TP4	I was able to carry out my work well with minimal time and effort. * §	0 (0)	1 (2.5)	7 (17.5)	20 (50.0)	12 (30.0)	2.15 (1.00)	0.55
TP5	I planned my work optimally. §	0 (0)	5 (12.5)	10 (25.0)	16 (40.0)	9 (22.5)	2.73 (0.96)	0.60

Table 1. Continued

Contextual performance (CP) scale		Response, n (%)				Mean (SD)	Item-to-scale correlation	
		0 (seldom)	1 (sometimes)	2 (regularly)	3 (often)			4 (always)
In the past 3 months...								
CP6	On my own initiative, I started new tasks when my old tasks were completed.	0 (0)	5 (12.5)	8 (20.0)	10 (25.0)	17 (42.5)	2.98 (1.07)	0.24
CP7	I took on challenging tasks when these were available. §	1 (2.5)	2 (5.0)	5 (12.5)	17 (42.5)	15 (37.5)	3.08 (0.97)	0.74
CP8	I worked on keeping my job-related knowledge up-to-date. *	1 (2.5)	2 (5.0)	11 (27.5)	14 (35.0)	12 (30.0)	2.85 (1.00)	0.59
CP9	I worked on keeping my work skills up-to-date. *	0 (0)	3 (7.5)	13 (32.5)	11 (27.5)	13 (32.5)	2.85 (0.98)	0.62

Table 1. Continued

Contextual performance (CP) scale		Response, n (%)				Mean (SD)	Item-to-scale correlation	
		0 (seldom)	1 (sometimes)	2 (regularly)	3 (often)			4 (always)
In the past 3 months...								
CP10	I came up with creative solutions for new problems.	1 (2.5)	3 (7.5)	11 (27.5)	18 (45.0)	7 (17.5)	2.68 (0.94)	0.52
CP11	I took on extra responsibilities.	1 (2.5)	2 (5.0)	6 (15.0)	23 (57.5)	8 (20.0)	2.88 (0.88)	0.60
CP12	I continually sought new challenges in my work.	0 (0)	4 (10.0)	10 (25.0)	15 (37.5)	11 (27.5)	2.83 (0.96)	0.70
CP13	I actively participated in meetings and/or consultations.	0 (0)	4 (10.0)	5 (12.5)	15 (37.5)	16 (40.0)	3.08 (0.97)	0.43

Table 1. Continued

Counterproductive work behavior (CWB) scale		Response, n (%)				Mean (SD)	Item-to-scale correlation	
		0 (seldom)	1 (sometimes)	2 (regularly)	3 (often)			4 (always)
In the past 3 months...								
CWB14	I complained about unimportant issues at work. * §	4 (10.0)	20 (50.0)	13 (32.5)	3 (7.5)	0 (0)	1.38 (0.77)	0.70
CWB15	I made problems at work bigger than they were.	15 (37.5)	20 (50.0)	4 (10.0)	1 (2.5)	0 (0)	0.78 (0.73)	0.73
CWB16	I focused on the negative aspects of a situation at work instead of the positive aspects. *	10 (25.0)	19 (47.5)	10 (25.0)	1 (2.5)	0 (0)	1.05 (0.78)	0.72

Table 1. Continued

		Response, n (%)				Mean (SD)	Item-to-scale correlation
		0 (seldom)	1 (sometimes)	2 (regularly)	3 (often)		
In the past 3 months...							
CWB17	I talked to colleagues about the negative aspects of my work. *	8 (20.0)	18 (45.0)	8 (20.0)	5 (12.5)	1 (2.5)	1.33 (1.02) 0.81
CWB18	I talked to people outside of the organization about the negative aspects of my work. *	10 (25.0)	18 (45.0)	7 (17.5)	4 (10.0)	1 (2.5)	1.20 (1.02) 0.76

\*Difficult to translate; †Adjusted after pilot-test.

*Pilot-test*

The pre-final version of the questionnaire was administered to 40 employees of Tufts Medical Center (n=18 men and n=22 women). On average, participants were 34.5 (9.8) years of age, and worked 45.9 (13.7) hours a week. See Table 2 for an overview of the sample descriptives.

Table 2. Descriptive statistics of the pilot-test sample

	Total (n=40)	Men (n=18)	Women (n=22)
Age in years, mean (SD)	34.5 (9.8)	35.2 (8.9)	33.8 (10.6)
Working hours/week, mean (SD)	45.9 (13.7)	50.8 (13.6)	41.8 (12.7)
Primary occupation, n (%)			
<i>Research position</i>	19 (47.5)	9 (50.0)	10 (45.5)
<i>ICT position</i>	1 (2.5)	1 (5.6)	0 (0)
<i>Education position</i>	2 (5.0)	0 (0)	2 (9.1)
<i>Clinical position</i>	12 (30.0)	7 (38.9)	5 (22.7)
<i>Administrative position</i>	6 (15.0)	1 (5.6)	5 (22.7)
Highest completed education level, n (%)			
<i>High school degree</i>	3 (7.5)	0 (0)	3 (13.6)
<i>College degree (e.g., Bachelor)</i>	12 (30.0)	4 (22.2)	8 (36.4)
<i>Master degree</i>	9 (22.5)	3 (16.7)	6 (27.3)
<i>MD degree</i>	15 (37.5)	10 (55.6)	5 (22.7)
<i>PhD degree</i>	1 (2.5)	1 (5.6)	0 (0)

Five participants (12.5%) mentioned that the **instructions** were not clear in terms of wording. Two of these participants were unsure what “how you carried out your work” meant, and three of these participants thought “conduct at work” was vague. One participant felt that the use of “conduct at work” had a negative interpretation, while two other participants said that it was not negative per se, but it were just words they would not normally use. During the translation process, conceptual issues had also arisen with these words. After deliberate discussion, the translators reached consensus on “how you carried out your work” as closest to the original meaning. As no better alternative was suggested during the pilot-test, and only a minority of participants reported an issue, it was chosen not to change the instructions.

Ten participants (25%) felt that the distinctions between the **answer categories** were unclear. This mainly concerned the distinction between “regularly” and “often,” with eight participants feeling that these categories are almost the same, and could also be placed the other way around. In addition, two participants felt that “seldom” and “sometimes” were almost the same. One participant felt that “seldom” should be worded as “rarely.” Finally, two participants wondered whether everyone would notice the change in answer categories for the CWB scale. Some participants suggested to rename the answer categories to “none of the time – some of the time – half of the time – most of the time – all of the time,” or to only name the extreme categories and number the middle categories. Another participant said that no matter how the answer categories are labeled, people will always have trouble distinguishing them, and they will be filled in like a VAS scale. As no clear alternative arose during the pilot-test, and only a minority of participants reported an issue, it was chosen not to change the answer categories in order to retain equivalence to the Dutch version.

Although participants stated that they had no major difficulties in understanding or answering most of the **items**, six items stood out during the pilot-test. Most comments were made on the items in the task performance scale. Twelve participants (30%) were unsure what was meant by “work result” in question **TP2** (“I kept in mind the work result I needed to achieve”). They made suggestions to change “work result” to for example “work goal,” “deadlines,” or “work outcome.” However, as there was no consensus on an alternative amongst the participants, nor amongst the expert committee, the question was not changed. Thirteen participants (32.5%) thought that question **TP3** (“I was able to distinguish main issues from side issues”) was oddly phrased. Most of the participants said they would never use the words “main issues and side issues,” and were unsure what they meant. Most of the participants suggested replacing these words with “prioritize.” Therefore, the question was changed to “I was able to set priorities.” Seventeen participants (42.5%) felt that question **TP4** (“I was able to carry out my work well with minimal time and effort”) had a negative interpretation. Participants felt you cannot carry out your work well with minimal time and effort. This means that you are lazy, and you take shortcuts. You need to put in time and effort to do your work well. Most participants suggested changing the question to whether you were able carry out your work “efficiently.” Therefore, the question was changed to “I was able to carry out my work efficiently.” Eleven participants (27.5%) indicated that question **TP5** (“I planned my work optimally”) was strangely phrased, although they understood what

was meant. Suggestions for rephrasing this question were diverse, including “I scheduled my work optimally,” “I planned my work efficiently,” “I managed my time well,” “I could get my work done in the best way possible,” and “with planning I was able to complete all my work tasks.” The translators and authors of the IWPQ agreed that the question should be changed to “I managed my time well.” In question **CP7** (“I took on challenging work tasks when these were available”), the word “these” was replaced by “they”, based on suggestion from three participants and the translators. Eleven participants (27.5%) commented on question **CWB14** (“I complained about unimportant issues at work”). Seven of these participants tripped over the word “unimportant,” and argued that if they complained about it, that meant the issue was not unimportant to them. Four of these participants wanted more specificity as to who and where they should have complained (e.g., to colleagues or to friends, at work or at home). Therefore, it was decided to change the question to “I complained about minor work-related issues at work.” In conclusion, a total of five items were changed based on the pilot-test results (marked in Table 1).

Almost all participants (85%) felt that all questions were applicable to their job. Two participants said that question TP5 (planning work optimally) was less relevant to their job, because as doctors, they had little influence on how many patients came in during the day, and which problems were presented. Two participants said CP8 (keeping job-related knowledge up-to-date) was less relevant to their job. Also, one participant said CP9 (keeping work skills up-to-date) was less relevant to their job. Furthermore, five participants reported reservations to answer the CWB questions honestly, because they felt the questions were a bit uncomfortable or intense to answer. Two participants said that the CWB questions were less relevant to them, one because solving problems (negative aspects) was a part of her work, and the other because she was not supposed to complain at work. Based on these few comments, it was not considered necessary to remove any questions from the questionnaire. Due to the generality of the questionnaire, it was considered inevitable that some questions are less relevant to some participants than others.

All participants (100%) stated that the completeness of the questionnaire was good. When asked, 16 participants (40%) had suggestions to expand the questionnaire to include all relevant aspects of their work performance. These suggestions mainly included determinants of individual work performance (e.g., job satisfaction, job tenure, and sleep quality), or indicators of individual work

performance that were previously included, but removed during the development of the questionnaire (e.g., relationship with co-workers and supervisor(s), collaboration with others, access to and use of supplies). Based on the suggestions, it was not considered necessary to add any new questions to the questionnaire. A short questionnaire with content identical to the Dutch version was considered most important.

### **Measurement properties of the pre-final questionnaire**

Descriptive statistics of the IWPQ items can be seen in Table 1, and descriptive statistics of the IWPQ scales can be seen in Table 3. Almost all items showed floor or ceiling effects (>15% at the lowest or highest answer category). At the scale level, the mean score for task performance was 2.79 (SD=0.69), 2.90 (SD=0.65) for contextual performance, and 1.15 (SD=0.73) for counterproductive work behavior. The mean scale scores are comparable to scores in The Netherlands, although the mean scale score for contextual performance was slightly higher than in The Netherlands (2.90 in the USA, versus 2.31 in The Netherlands) [9]. There were no ceiling or floor effects on the scale level. Five percent of the participants showed the highest score (4, “always”) for the task performance scale, and the contextual performance scale. Five percent also showed the lowest score (0, “never”) for the counterproductive work behavior scale.

The Cronbach’s alpha for the task performance, contextual performance, and counterproductive work behavior scales were 0.79, 0.83 and 0.89, respectively (Table 3). The item-to-scale correlations were sufficiently high ( $r > 0.40$ ), except for item CP6 (“On my own initiative, I started new tasks when my old tasks were completed”), which correlated  $r = 0.24$  on the contextual performance scale (see Table 1).

Based on the cultural adaptation process, and the comments provided by the participants of the pilot-test, the content validity of the American-English IWPQ was judged to be good. Almost all participants in the pilot-test considered the questions to be applicable and relevant to their job, and all participants felt that the completeness of the questionnaire was good.

Table 3. Descriptive statistics of the pre-final American-English IWPQ scales

	Range (0 - 4)	Mean (SD)	Median	% floor- effects (score 0)	% ceiling- effects (score 4)	Cron- bach's $\alpha$
Task performance	1.40 - 4	2.79 (0.69)	2.90	0	5	0.79
Contextual performance	1.50 - 4	2.90 (0.65)	2.88	0	5	0.83
Counter- productive work behavior	0 - 3.20	1.15 (0.73)	1.10	5	0	0.89

## Discussion

The goal of the current study was to cross-culturally adapt the Individual Work Performance Questionnaire (IWPQ) from the Dutch to the American-English language, and to assess the questionnaire's internal consistency and content validity in the American-English context. The cross-cultural adaptation was systematically performed, resulting in an American-English version of the IWPQ that equals the original version. In general, participants were positive on the comprehensibility of the questionnaire. A few changes were made to optimize the comprehensibility of the questionnaire. Here, the consideration of not changing the wording of a question in order to keep it similar to the original question, versus changing the wording of a question in order to obtain conceptual equivalence to the original question, is important. For example, the answer category labels of the IWPQ were not changed, in order to retain equivalence to the Dutch version, and because no alternative arose that was believed to improve comprehensibility. On the contrary, the wording of task performance items 3 (*"I was able to distinguish main issues from side issues"*) and 4 (*"I was able to carry out my work well with minimal time and effort"*) was changed in order to improve comprehensibility. In Dutch, it was chosen to give a description of "prioritizing" and "efficiently," as these words are hardly ever used directly. However, based on American participants' suggestions to improve comprehensibility, these items were shortened to more directly ask for "prioritizing" and "working efficiently."

All participants were positive on the completeness of the questionnaire, and almost all participants indicated that all the questions were relevant and

applicable to them. This indicates good content validity of the questionnaire. Thus, there appear to be no cultural differences between The Netherlands and America in measuring the concept of individual work performance, and the indicators used to measure the concept of individual work performance seem to be equivalent over these contexts. Although additional indicators of individual work performance suggested by participants in the pilot-test (e.g., relationship with co-workers and supervisor(s), collaboration with others, access to and use of supplies) might have been included when developing the IWPQ from scratch in America, a short questionnaire with identical content to the Dutch IWPQ was considered most important in the current study. The generalizability of the questions in the Dutch IWPQ was probably promoted by the fact that people from multiple countries (including the USA) were involved in the developmental stages of the IWPQ, for example, during item generation [18].

If the IWPQ items have kept the same meaning after the translation, the American-English questionnaire is expected to retain the same factor structure as in The Netherlands. The sample size in the current pilot-test ( $n=40$ ) was too small to conduct a confirmatory factor analysis. De Vet et al. [16] recommend a sample size of at least  $n=100$  to perform a reliable factor analysis. However, item-to-scale correlations were examined, and were similar to the item-to-scale correlations in The Netherlands. All items loaded sufficiently high on the expected scales, except for item CP6 (*“On my own initiative, I started new tasks when my old tasks were completed”*). The low loading of this item on the contextual performance scale suggests that this item has a different meaning in the USA than in The Netherlands, either due to the translation, or due to cultural differences. However, no specific comments were made regarding this question during the pilot-testing phase, so the reason for the low loading is unclear. Future research should administer the American-English IWPQ in a larger sample (at least  $n=100$ ), so that its factor structure can be examined, and the loading of the items on each scale can be examined in more detail.

The measurement properties of the Dutch and American-English IWPQ appear to be similar. The mean item and scale scores appear to be similar in both versions, although the mean scale score for contextual performance was slightly higher for the American-English than Dutch IWPQ. Nevertheless, there were no considerable ceiling or floor effects at the scale level. The internal consistencies of the American-English IWPQ task performance, contextual performance, and counterproductive work behavior scales were 0.79, 0.83, and 0.89, respectively. This

is similar to the Dutch version, where the scale reliabilities are 0.78, 0.85, and 0.79, respectively. The internal consistency of the American-English CWB scale is higher than in The Netherlands.

### **Limitations**

A limitation of the current study was that participants were aware that the questionnaire measured individual work performance, due to the informed consent procedure before the study. Secondly, in the current study, a researcher was sitting next to the participants while they were filling in the questionnaire. Finally, some participants reported reservations to answer the CWB questions honestly, because they felt the questions were a bit uncomfortable or intense to answer. All these factors may have elicited socially desirable answers, and resulted in different scores on the American-English version than the Dutch version of the questionnaire. In general, we recommend leaving out the questionnaire title and scale names when administering the questionnaire, so that participants are less aware they are filling in a questionnaire on individual work performance. We also recommend that participants' answers are always anonymous and are treated confidentially. It should be guaranteed that only group level outcomes will be reported to managers or companies, obtained in large enough groups, so that results can never be traced back to individual participants.

The pilot-test in the current study was conducted in a relatively high-educated sample, with participants primarily working in a pink or white collar job. This may limit generalizability of the results to lower-educated workers, and blue collar workers. Although, in general, the translators were positive on the questionnaire's comprehensibility, applicability, and completeness for lower-educated workers, and blue collar workers, one translator had concerns about the use of the word "priorities" in these groups. Ideally, the comprehensibility, applicability, and completeness of the American-English IWPQ, as well as its internal consistency and content validity, should still be examined in these groups.

### **Future research**

Although the pilot-test results indicate good internal consistency and content validity of the American-English IWPQ, it is only after the cross-cultural translation and adaptation that the real cross-cultural validation takes place [16]. In a larger and more heterogeneous sample, special attention should be paid to the measurement invariance of the questionnaire in the original and the new target population.

Measurement invariance means that a measurement instrument, a scale, or an item, functions in exactly the same way in different populations [16]. This can be examined, for example, using factor analysis or item response theory (IRT) techniques. Future research should perform confirmatory factor analysis in a larger and more heterogeneous sample, and examine if (and if so, why) item CP6 loads insufficiently high on its original dimension. IRT techniques are also a powerful method with which to detect differential item functioning (DIF), by comparing the item characteristic curves of the items in the original version and the translated version [16]. This can give insight into whether the difficulty of an item has changed in the original and translated version. Future research should also further examine the reliability, construct validity, and responsiveness of the American-English IWPQ.

### **Conclusion**

The cross-cultural translation and adaptation of the IWPQ from the Dutch to the American-English language was conducted without major difficulties. The comprehensibility, applicability, and completeness of the translated version of the IWPQ was considered to be good. Also, its internal consistency and content validity appeared to be good. The translated questionnaire can now be used to measure, for example, the effectiveness of workplace interventions on individual work performance in an American-English speaking context. Future research should further examine the measurement invariance, reliability, validity, and responsiveness of the American-English IWPQ in a larger and more heterogeneous sample.

### *Thank you*

We wish to thank Kimi Uegaki, Tammy Rubinstein, Fenna Leijten and Nico Pronk for their help in translating the questionnaire, the EMGO+ Institute for Health and Care Research for providing the travel grant, Tufts Medical Center for their hospitality, and the Tufts Medical Center employees for their participation in the pilot-test.

## References

1. European Commission. EUROPE 2020: A strategy for smart, sustainable and inclusive growth. 2010. Available from: <http://ec.europa.eu/europe2020/>.
2. Campbell JP. Modeling the performance prediction problem in industrial and organizational psychology. In: Dunnette MD, Hough LM, editors. *Handbook of industrial and organizational psychology, Vol.1 (2nd ed.)*. Palo Alto, CA, US: Consulting Psychologists Press; 1990. p. 687-732.
3. Austin JT, Villanova P. The criterion problem: 1917-1992. *Journal of Applied Psychology* 1992;77(6):836-874.
4. 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.
5. 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.
6. Viswesvaran C, Ones DS. Perspectives on Models of Job Performance. *International Journal of Selection and Assessment* 2000;8(4):216-226.
7. Borman WC, Motowidlo SJ. Expanding the criterion domain to include elements of contextual performance. In: Schmitt N, Borman WC, editors. *Personnel Selection in Organizations* San Francisco, CA: Jossey Bass; 1993. p. 71-98.
8. 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.
9. 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).
10. Dalal RS. A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *J Appl Psychol* 2005;90:1241-1255.
11. 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.

12. Podsakoff PM, MacKenzie SB. A second generation measure of organizational citizenship behavior. Indiana University, Bloomington; 1989.
13. 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.
14. Lerner D, Amick BC, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. *Medical Care* 2001;39(1):72-85.
15. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25(24):3186-3191.
16. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine*. Cambridge University Press; 2011.
17. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297-333.
18. 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).