Danish translation and validation of the Occupational Balance Questionnaire

Hansen, Alice Ørts; Boll, Mette; Skaarup, Line; Hansen, Tina; Dür, Mona; Stamm, Tanja; Kristensen, Hanne Kaae

Published in:
Scandinavian Journal of Occupational Therapy

DOI:
10.1080/11038128.2020.1842491

Publication date:
2022

Document version:
Accepted manuscript

Citation for published version (APA):

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying this open access version.

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 15. Sep. 2023
Danish translation and validation of the Occupational Balance Questionnaire

Running heading

Validation of the Danish OB-Quest

Authors

Alice Ørts Hansen1,2, Mette Boll3, Line Skaarup3, Tina Hansen4, Mona Dür5, Tanja Stamm6, Hanne Kaae Kristensen2,3

1) Department of Rehabilitation, Odense University Hospital, Odense, Denmark
2) Department of Clinical Research, University of Southern Denmark, Odense, Denmark
3) Health Sciences Research Centre, UCL University College, Odense, Denmark
4) Department of Occupational Therapy, Faculty of Health, University College, Copenhagen, Denmark
5) Department of Health Sciences, Master Degree Programme Applied Health Sciences, IMC University of Applied Sciences, Krems, Austria
6) Section for Outcomes Research, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria

Corresponding author

Alice Ørts Hansen
Abstract

Background: Occupational balance is a key concept in occupational science and occupational therapy. However, it is not well operationalised and instruments to evaluate occupational balance are scarce.

Aim: To translate and validate a Danish version of the 10-item Occupational Balance Questionnaire (OB-Quest).

Material and Methods: Translation was conducted using the dual-panel approach. Thereafter, data from 262 healthy Danish students were analysed regarding fit to the Rasch model.

Results: The translation of the questionnaire to Danish required only minor cultural adaptations. Ninety percent of the participants were female. The mean age (+/-SD) was 26.10 (7.05) years. The reliability (person separation index) was questionable (0.63), and the scale showed multidimensionality. Two items (1 and 9) showed misfit to the Rasch model. Differential item function by gender was detected in one item (item 4). After deleting items 1 and 9 and splitting item 4 into two gender-specific items, the new 9-item scale showed good overall and individual item fit. However, reliability remained low (0.59) and some elements of the latent variable (occupational balance) were not sufficiently represented.

Conclusions and Significance: Further development and testing of the Danish OB-Quest is needed before implementation in clinical practice or research involving healthy subjects.

Keywords

Occupational science, occupational therapy, Rasch model
Introduction

Occupational balance (OB) is a fundamental concept in occupational science and occupational therapy (OT) and is often associated with health and well-being (1-4). OB has been defined as the subjective ‘perception of having the right amount and variation of occupations’ (5). However, there is no agreed or shared definition of OB. In occupational science literature, shows that the concept has been diversely understood and used in both practice and in research (5-8). Westhorp (6) synthesized OB from three perspectives: balance, occupation and epidemiology. She suggested that achieving OB is a subjective experience and that the individual person decide, choose and control the range of occupation, and the amounts of time, energy, and other resources expended in order to achieve engagement in occupation. This subjectivity can be seen in how individuals aim for balance on a daily basis in relation to environmental demands, skills development and motivation.

According to Westhorp, balance in occupation is a product of having the opportunity to engage and develop one’s capabilities in meaningful occupations and having the right amount of experience of flow, along with an understanding of roles and occupations. Another element of OB is individual appreciation of the effects of occupations and interaction with social context. Finally, balance is seen as dynamic and dependent on experiences of control over actions and on receiving recognition (6). Emphasizing flow theory, Jonsson and Persson (9) developed an experiential model of dynamic balance in occupation, on three dimensions: “exacting” – covering experiences of everyday occupation that exceed the person’s actual skills, “flowing” – encompassing experiences with a reasonable match between skills and challenges, and “calming” – comprising the low challenge experiences of relaxation, boredom or apathy. Jonsson and Persson argued that, when these dimensions of challenges and skills in occupations are reasonably balanced, they constitute a dynamic and healthy occupational experience pattern. This may contribute to development, well-being and meaning for the person, and contribute to interaction with one’s social network and societal context (9, 10).
OT seeks to engage people in meaningful occupations (11) and OB is often evaluated in the context of occupational roles, patterns, well-being and health. From a preventive and health-promoting perspective, the “Well Elderly Study” showed that meaningful occupations in a satisfying pattern had a positive effect on health and well-being (12). In several studies by Wagman and Håkansson et al. (1, 4, 13-15) associations were found between individual OB and subjective health, i.e. self-rated health, satisfaction and well-being. Health indicators, i.e. manageability, personally meaningful occupations, and OB needed to be combined to prevent stress and sick leave, e.g., among women of working age (13). In rehabilitation, reduced functioning and changed living conditions have been seen to affect occupational patterns. Furthermore, Forhan and Backman (16) also found that experience of health was associated with OB for people with chronic illness, occupational limitations and participation restrictions. Their study found that OB was achieved as a result of experiencing a combination of enjoyable occupations as part of occupational patterns and feeling able to manage life with a chronic illness. In line with these studies, a recent study by Yazdan et al. found that the concept of meaningful engagement in relation to balancing meaningful occupations, needs and priorities was emphasized by OTs in connection with OB. Moreover, OB was associated with occupational roles, and participation. Hence, OTs also related OB to purposeful participation in occupations, in that OB was seen to provide structure in life, a sense of purpose, competency and satisfaction (8).

OT aims to enable people to engage in personally meaningful occupations and support them with strategies and tools to achieve subjective and dynamic management of their everyday lives, health and well-being (11, 17, 18). Thus, OB is a relevant concept for OT practice and research, from not only an illness-preventive and health-promoting perspective but also from a rehabilitation perspective. In this context, it is important to have valid and reliable OB measurement instruments.
for the clinical practice of health promotion and rehabilitation services. Therefore, there has been an ongoing focus on exploring OB in the past decades (5-7) and questionnaires have been developed and used in occupational science and occupational therapy research among both healthy adults and people living with illness and disabilities (4, 7, 19, 20).

The use of self-reported outcome instruments in Danish social and health care and in research has become increasingly important, because user perspectives are considered essential, in terms of goal attainments and other outcomes, e.g., of OT interventions. A user perspective of OB was explored in 2014 by Dür et al., (7) with the aim of identifying components of OB. Both healthy people and people who suffered from various rheumatic conditions took part. Based on the participants’ life stories, eight user-identified components of OB were developed: 1. Challenging and relaxing activities, 2. Activities with acknowledgement by the individual and by the sociocultural context, 3 Impact of own health condition on activities, 4. Involvement in stressful activities and fewer stressing activities, 5. Satisfaction with the amount of rest and sleep, 6. Engagement in a variety of activities, 7. Adaptation of activities according to changed living conditions, and 8. The meaning of activities intended to care for oneself and for others (7). Hence, in this study OB was understood as a complex concept, however, no definitions were generated (7). OB was seen to change dynamically according to individual experience of the number of different categories of occupation. It also changes with the experience of being in control of the adaptation to changed living conditions. The amount of variation in occupation, along with categories and the meaning of changes to living conditions lead to a perception of OB as a dynamic concept. From these findings, the self-reported Occupational Balance Questionnaire (OB-Quest) was developed in 2014 by Dür et al. (7). OB-Quest is currently available in English and German (7). In the evaluation of a construct of interest, such as OB, and to achieve effectiveness in relevant interventions, it is desirable that an
outcome measure contains a unidimensional abstraction of the construct and its magnitude (21). A literature search of studies that have validated the 10-item OB-Quest as a measuring instrument established that no study has evaluated the 10-item version of the scale for unidimensionality or by using the Rasch model. Given that our aim is to use OB-Quest in education, prevention and health promotion, along with rehabilitation practice and research, it is relevant to translate it to Danish and test the psychometric properties. From a prevention and health promotion perspective, we have prioritized first the validation of OB-Quest in a healthy population. Based on the components of OB identified by Dür et al., the aim of the current study was to develop a Danish translation of the self-reported 10-item OB-Quest and to determine its validity and reliability when used among a group of healthy adults in a Danish context.

Materials and methods

Study design
This study consisted of two phases. In phase I, the study design involved translation and cross-cultural adaptation of the OB-Quest into Danish, and phase II comprised a survey to test the psychometric properties of OB-Quest/DK.

Phase I: Translation and cross-cultural adaptation
Translation and cross-cultural adaptation of the OB-Quest into Danish were conducted in accordance with the guidelines of the dual-panel approach (22). First, the English version of OB-Quest was translated individually by a panel consisting of five bilingual participants with good English language skills, one of whom had English as their native language and one of whom was a professional translator. The panel then met to achieve consensus on a first version. A lay panel then discussed this first version one item at a time, including format, meaning and wording, and reached
a consensus on a second version. The lay panel consisted of seven lay people purposefully selected so they varied in age (mean: 44 years [range 17-78 years]), gender (women $n = 4$, men $n = 3$) and educational background (lower secondary school – university degree). As recommended in the dual-panel approach guidelines, none had a long-term higher education (22). The first, third and last authors participated in both panels to ensure that the content of the original OB-Quest was retained. To ensure optimal understanding of the items, the content of each item was discussed among the authors – including the developers of the original questionnaire (4th and 5th authors) during the translation process. Throughout the translation process, disagreements were resolved by consensus and finally settled by the first and last authors.

The linguistic content, comprehensibility and feasibility of the OB-Quest/DK were assessed by cognitive interviews with 18 OT students. The participants completed the electronic OB-Quest and were subsequently interviewed about how they understood the items and whether they had difficulty answering any of the items. Any difficulty in understanding an item was recorded and discussed among the authors. Adjustments were made when needed before the psychometric testing of the OB-Quest/DK.

**Phase II: Psychometric testing**

**Participants**

The first test groups consisted of a group of young Danish students, which was in line with international and national health prevention initiatives that focus on reduction of stress especially among healthcare professionals (23-26). Students enrolled at Danish University Colleges, whether as nursing, OT or multimedia students during the recruitment period from 01.02.19 to 28.02.19 were invited to participate in the psychometric testing. None was excluded. The participants were
invited from a range of educational institutions and regions of Denmark to reflect different
demographic factors, such as gender, means of transport, living in an urban or rural area, etc.

Measures

Occupational Balance Questionnaire (OB-Quest)

The earliest version of the original OB-Quest contained seven items, which were derived from
qualitative interviews of patients and healthy individuals (7). The OB-Quest was developed to be a
generic questionnaire and its psychometric properties were explored among 76 healthy people above
median age 38 years and 132 people with rheumatoid arthritis, median age 59 years, and 43 with
systemic lupus, median age 46 years. The items reflect seven of the eight user-identified
components of OB. No item reflected the component meaning of activities intended to care for
oneself and for others, since not everyone cares for others. Based on analysis by the Rasch model,
the developers proposed that the components ‘challenging and relaxing activities’, ‘satisfaction with
the amount of rest and sleep’, and ‘adaptation of activities according to changed living conditions’
should be divided into two items each. This resulted in the final OB-Quest, with ten items, which
was translated into Danish and further tested in this study (7) See all items in Table I. Each item is
rated on a scale with three response categories, where a score of 1 indicates a high degree of OB
and a score of 3 indicates a low degree of OB. As such, higher scores on the OB-Quest represent
lower OB (7).

Demographics

In the psychometric testing, the respondents were asked to answer a number of closed questions in
addition to the OB-Quest/DK, relating to: gender (male/female), age (from 18 to > 60 years), civil
status (married, in a relationship, single, divorced), currently enrolled on a bachelor degree
programme (OT, multimedia, or nursing), current semester (1-7), work in addition to study (yes/no),
time used for transportation to and from education site (min) and with children living at home
(yes/no).

**Procedure of the data collection for the analysis by the Rasch model**

The Danish version of the 10-item OB-Quest/DK was administered to the respondents using
SurveyXact 2019. A teacher at each educational institution was engaged as gatekeeper. The link to
the survey was shared by the teacher during a lecture and the students had time to fill in the
questionnaire during the lecture or at a later, self-chosen, opportunity, as the link to the survey was
open for one month. The fourth author was in contact with the gatekeepers during the data
collection period to encourage participation and offer information on the study. It took 3-10 minutes
to complete the electronic OB-Quest/DK.

**Analysis by the Rasch model**

The Rasch measurement model is used to explore unidimensionality of measurement instruments,
which is one of several measurement properties (27). Analyses with a Rasch model were applied in
the development of the original German OB-Quest (7). Analysis by the Rasch model implies formal
testing of a rating scale against a mathematical measurement model that operationalizes formal
measurement. An important principle of the Rasch model is specific objectivity. This implies
invariance, which means that the comparison between two persons should be independent of the
items that have been used and vice versa; therefore, the instrument should work in the same manner
across all persons and items (27).
For the analysis by the Rasch model, RUMM2030 software was used (28). In the case of OB-Quest/DK, the Rasch model specifies that the probability of a response of 1, 2 or 3 is a logistic function of the difference between the respondent’s degree of the measured latent variable (OB) and the degree of the latent variable represented by the item. Logits (log-odd units) constitute the unit of measurement to report the relative differences between the estimates of a person’s degree of the latent variable and item difficulties and are an equal interval level of measurement. Persons (i.e., respondents) and items are located on the same measurement scale with the mean item location set at zero logits (21,27, 29). Accordingly, the ordinal scores from the OB-Quest/DK items are expressed as linear measures, where negative values reflect easy items and higher degree of OB, while positive values reflect difficult items and lower OB (27, 29, 30). However, the data obtained by the OB-Quest/DK must adhere to certain requirements articulated by the Rasch model, before it can be considered to measure a continuous latent variable of less or more OB (27, 29, 30), namely:

- Unidimensionality: all items in the rating scale measure the same single construct.
- Monotonicity: the scale items function hierarchically from easy to difficult, with increasing item scores corresponding to increasing degree of the measured latent variable.
- Homogeneity: The ordering of the items from easy to difficult is the same for all respondents.
- Local independency: a score on one item does not depend on another item’s score.
- Absence of differential item function (DIF): an item’s score does not differ due to other factors, such as age or gender for persons with an equal degree of the measured latent variable.

Whether the OB-Quest/DK accomplishes these requirements was investigated through fit of the data to the Rasch model, and the analysis followed recommended procedures (21, 27, 29, 31). Fit to the Rasch model was examined statistically and graphically. This was carried out for items, persons,
and for different groups (e.g., class intervals) according to their locations on the measured variable. Ideally, the class intervals should be approximately equally distributed with around 50 persons in each (32).

Statistically, model fit was examined using standardized fit residual values, which express the differences between observed responses on the OB-Quest/DK and those expected by the model, and by analysing them by means of chi-squared ($\chi^2$) statistics and analysis of variance (ANOVA) of the residuals across class intervals (21, 29). Fit residual values between ± 2.5 are regarded as acceptable (21, 29). High item fit residuals are signs of under-discrimination and might reflect multidimensionality, while low residual values indicate over-discrimination and might reflect potential redundancy or item dependency within the item set (32). Chi-squared statistics and ANOVA should indicate non-significant (Bonferroni adjusted) deviations from model expectations (21, 29).

Item fit was also examined graphically, by plotting the observed item responses for each class interval against the model expectations, which are displayed as an item characteristic curve (ICC) (32). Based on the statistical analysis, overall model fit is provided in RUMM2030. Model fit is present when a summary fit residual statistics for items and persons approach a standardized mean value of zero and an SD of 1.0, and a summary item $\chi^2$ statistic is non-significant ($p > 0.05$) reflecting invariance of the items across the different class intervals (21, 29, 32). Missfitting items or persons can be removed to try to improve overall model fit.

The OB-Quest/DK consists of polytomous items with three response categories that are ordered to reflect a movement towards lower degree of OB (7). The boundaries between adjacent categories are called thresholds (27, 33). As the number of thresholds is one less than the number of response categories, there are two thresholds for each of the OB-Quest/DK items. Thresholds reflect
positions on the latent variable where either of the adjacent responses is equally probable (27, 33). For fit to the Rasch model, monotonicity is expected and implies that the transition from one score to the next is consistent with the increase in the underlying variable (i.e. ordered thresholds) (27, 33). Threshold ordering was examined using a threshold map and category probability curves. The presence of disordered thresholds might suggest that the response categories do not work as intended, either due to multidimensionality, too many response options or unclear item and category descriptions, and this may be resolved by combining adjacent categories (33).

Local independence was investigated using a residual correlation matrix of the items. Local item dependence (LID) was evident by item residual correlations above 0.2, reflecting that the entire correlation between the items is not captured by the latent trait (34). This might happen when the content of a previous item affects responses to a subsequent (dependent) item. This can be dealt with by grouping local dependent items into a “super-item” to absorb the impact of LID (35).

Differential item functioning (DIF) refers to item bias that occurs when subgroups with a similar degree of the measured variable have a different response pattern to any item (36). The analysis uses a 2-way ANOVA on the residuals for each item across the subgroups and across the class intervals. DIF can occur as either uniform DIF, where item responses differ uniformly across the measured variable (i.e. a main effect), or as non-uniform DIF, where differences in item responses between subgroups vary across the measured variable (i.e. an interaction effect). The Bonferroni correction was used to adjust for multiple testing, keeping the Type 1 error to 5%. Uniform DIF can be dealt with by splitting the item into group-specific items. Non-uniform DIF is usually removed as it reflects misfit to the model across the continuum (29). In the current study, DIF was examined by gender, age, bachelor degree programme currently enrolled on, time in education, work in addition to study, and time used for transportation to education site. Table II presents the subgroups
and how they were created to ensure sufficient numbers in each class interval and subgroups for the 2-way ANOVA. However, this was not possible for the gender subgroup.

As a part of the overall model fit, reliability and unidimensionality of the scale are reported.

Reliability was examined using Cronbach’s alpha and the Person Separation Index (PSI), the Rasch equivalent of Cronbach’s alpha, except that it is calculated from the logit scale person estimates (27). The PSI indicates the power of the latent variable to discriminate amongst persons. There are various different reports about the acceptable values of these reliability indices, ranging from 0.70 to 0.95 (37). It has been suggested that $\alpha \geq 0.90 = \text{excellent}$, $0.90 > \alpha \geq 0.80 = \text{good}$, $0.8 > \alpha \geq 0.7 = \text{acceptable}$, $0.7 > \alpha \geq 0.6 = \text{questionable}$, $0.6 > \alpha \geq 0.5 = \text{poor}$, $\alpha < 0.5 = \text{unacceptable}$ (37, 38). In the RUMM2030 software, the PSI also reflects the power of the fit statistics, which can be excellent, good, reasonable, low, or too low. If the PSI is below 0.70, then the fit statistics that are obtained may not be reliable because of too large error variance (32).

Unidimensionality of the OB-Quest/DK scale is defined as the absence of any meaningful pattern in the residuals, which was assessed by Principal Component Analysis (PCA) (32, 39). The loading between items and the first residual factor were examined. This pattern was used to define two subsets of items, consisting of items with positive and negative loadings, respectively (32, 39). The difference of the person location estimates for each person from these two subsets of items was investigated using a series of t-tests (40). Unidimensionality was inferred if less than 5% of the sample from the two subsets showed a significant difference in person location estimates (40), or if the value of 5% fell within an exact binomial 95% confidence interval (CI) of proportions (27, 39).

After taking together the overall model fit and the fit statistics at item level and applying strategies to improve model fit in an iterative process, the scale to sample targeting of the final OB-Quest/DK scale was evaluated. Targeting is defined as the extent to which the range of the measure matches
the range of the measure in the study sample (27). A well-targeted rating scale would have both item and person mean locations of around zero and there would be enough items of varied difficulty (i.e. measuring varying degree of OB) to match the spread of scores among respondents (29, 32). If the OB-Quest scale is poorly targeted, i.e. too easy or too difficult, respondents may report a high degree of OB (floor effect) or a low degree of OB (ceiling effect). To assess targeting, we examined the person-item thresholds distribution map, which visually depicts person locations against item locations (32).

It is suggested that a sample size of 250-500 will provide a good balance for the statistical interpretation of the fit statistics (41, 42). Therefore, the sample size of 262 respondents in the present study was considered adequate. In RUMM2030, likelihood ratio statistics is applied to decide which polytomous version of the model to use (32). The ‘Partial credit model’ (PCM) allows the ordered response options to be free to vary in structure from item to item (27), while the ‘Rating scale model’ (RSM) assumes that the distance between the thresholds is equal across items (27). A significant likelihood ratio test ($\chi^2$ (df) of 69.54) (8), (P< 0.001) indicated that the unrestricted PCM should be used. In RUMM2030, the estimation of the item and person parameters is based on a pairwise conditional maximum likelihood algorithm, which enhances the stability and robustness of the estimates, especially when there might be relatively few cases in some categories for some items (27).

**Ethics**
Participants were introduced orally to the project before following the link to the survey, and in written detail in the electronic survey. Prior to participation, all participants signed a consent form. The project was accepted by the Danish Data Protection Agency (no. 2015-57-0016-022). Due to the nature of the study, approval by the Research Ethics Committee was not required, in accordance with Danish legislation on research ethics. Permission to translate the OB-Quest was obtained from
the developers (fourth and fifth authors) and it was translated in collaboration with the copyright holder.

Results

Translation and cross-cultural adaptation

In the translation process, both the English and the German versions of OB-Quest were consulted to ensure validity, and several issues were discussed among the authors to maintain the understanding of the concept and the framing of the questions. Regarding item 3 “Do you generally receive enough appreciation for activities in your everyday life?” it was discussed whether the purpose was to assess if the person feels appreciated or to assess how much appreciation the person received. Due to this discussion, the answers were changed to “I receive enough appreciation”, “I receive some appreciation” and “I do not receive any appreciation” instead of “I receive quite a lot of appreciation”, “I receive enough appreciation” and “I do not receive any appreciation”. This was because the purpose was to assess if the person gets sufficient appreciation, rather than the amount of appreciation received. Regarding items 5 and 9, it was specified that health includes both physical and mental health. Concerning item 10, it was agreed by the developers that changes in professional life or employment also include education. Furthermore, to ensure the intended understanding of activities, an explanation is given in parentheses, both in the introduction and when the term is used in an item. Regarding the use of the word stress, it was pointed out in item 4 that it concerns the subjective experience of stress and not physical symptoms of stress. Cognitive testing of the pre-final version revealed no major issues regarding completing the questionnaire; only minor comments about the layout and linguistic content were made.

Psychometric testing
In total, 263 out of the 423 distributed questionnaires were completed (62%). However, one questionnaire had more than 50% missing values and was therefore excluded. Accordingly, 262 respondents with full data records were included in the analysis (no missing values).

Table II displays demographics of the sample and the grouping of the person factors used in the analysis of DIF. Fifty percent of the sample attended the bachelor’s degree programme in OT, and about 90% were females. The mean age (SD) was 26.10 (7.05) years.

Item analysis by the Rasch model of the OB-Quest/DK

Table III (Analysis 1) shows that the initial analysis of all 10 items resulted in a significant item-trait interaction (chi-square (df) = 74.12 (40), p < 0.001) and a residual mean value (SD) for items of 0.08 (1.27), which indicates some degree of misfit between the data and the Rasch model. The residual mean value (SD) for persons was -0.19 (1.06), indicating no serious misfit among the respondents in the sample. The unidimensionality t-tests resulted in 8.1% (95% CI: 5.4;10.7) statistically significant different person estimates, based on the two most divergent subsets of items within the OB-Quest/DK scale; suggesting multidimensionality. The PSI was 0.63 and Cronbach’s alpha was 0.65, indicating a reasonable power of analysis of fit and questionable reliability, which implies that the translated set of items did not discriminate well between persons with different degrees of OB.

Table IV displays the fit statistics at item level. All items had ordered thresholds, indicating that the responses to the items are consistent with the metric estimate of the underlying construct of OB.
Item 1 (Activities under-demanding) showed misfit with a positive fit residual of 2.6 and significant fit statistics (P < 0.005 (Bonferroni adjusted)). Figure 1a shows that this item forms a flatter curve than the expected scores, which indicates that this item is under-discriminating, i.e., reflects multidimensionality. Item 9 (Adapt activity health) and item 10 (Adapt activity living condition) showed misfits in terms of significant fit statistics (P < 0.005 (Bonferroni adjusted)). Item 9 appears to be the more problematic item of those two, as it displays a medium high negative fit residual and highly significant fit statistics (p<0.001). Figure 1b shows that this item forms a steeper curve than the expected scores. This means that item 9 is over-discriminating, which may indicate potential redundancy or dependency within the item set. The results of the 2-way ANOVA of the standardised residuals showed uniform DIF by education for item 1 and by gender for item 4, based on the Bonferroni-adjusted value of 0.002. This is displayed in Figure 2. For the other person factors, no uniform nor non-uniform DIF was present. LID reflected by residual correlations > 0.20 was found for item pair 6 and 7 (r = 0.26) and for item pair 9 and 10 (r = 0.34).

To obtain a satisfactory overall model-fit, an iterative process with various different strategies was carried out, and the best-fitting solution was as follows. Item 1 was regarded to be most misfitting and was removed from the scale. As seen in Table III (Analysis 2), this resulted in improved overall fit statistics with a non-significant item-trait interaction (chi-square (df) = 38.14 (37), p = 0.076) and a residual mean value (SD) for items of -0.01 (1.07). The lower CI band of 4.3% for the t-tests suggest unidimensionality. Although the reliability increased slightly, with a PSI of 0.65 and Cronbach’s alpha of 0.67, these magnitudes still indicate questionable reliability. At item level,
some problems seen in the initial analysis presented in Table IV were resolved, in terms of adequate
fit statistics for item 10, and a slightly reduced LID for item pair 6 and 7 of (r=0.21). However, item
9 persisted in showing misfit in terms of significant fit statistics (P < 0.005 (Bonferroni adjusted))
and high residual correlation with item 10 (r=0.33). Item 4 continued to show uniform DIF by
genre, as displayed in Figure 2.

Item 9 was then regarded as the most misfitting item and was removed from the new 9-item scale.
Table III (analysis 3) shows that this resulted in markedly improved overall fit statistics with a non-
significant item-trait interaction (chi-square (df) = 28.07 (34), p = 0.257) and a residual mean value
(SD) for items of 0.09 (0.87). The t-tests suggest unidimensionality with 2.7% (95% CI: 0.0;5.3)
statistically significant different person estimates, based on the two most divergent subsets of items
within the OB-Quest/DK scale. The reliability decreased slightly with a PSI of 0.59 and Cronbach’s
alpha of 0.62, reflecting poor to questionable reliability. As shown in Table V, there were no
misfitting items and the LID for item pair 6 and 7 was resolved. Item 4 persisted with uniform DIF
by gender, as displayed in Figure 2. Although the size of the male sample of 25 might be too small,
the DIF was resolved by splitting the item into two gender-specific items. This further improved the
item-trait interaction (chi-square (df) = 29.38 (37), p = 0.347) and increased the PSI to 0.60.

Since the results from analysis 4 might be unreliable due to the small male sample size, targeting of
the OB-Quest/DK scale was assessed during both fitting solutions in analyses 3 and 4. Figure 3
displays the targeting and shows that the OB-Quest/DK scale presented reasonable targeting, with a
match between overall spread of items and spread of respondents. No respondents gave a score of 3
(high degree of OB) for all items across the scale (0% ceiling effect) and only one respondent gave
a score of 1 (low degree of OB) for all items (0.4% floor effect). However, there were many gaps
along the item-thresholds continuum (between logit -3 to -2 and between logit -1 to 0). This might indicate that the OB-Quest/DK scale is not able to detect small changes in respondents across the whole continuum of OB.

Discussion

This study aimed to determine the validity and reliability of the Danish version of the 10-item OB-Quest/DK. Our initial analysis of all ten items showed overall misfit to the Rasch model, lack of unidimensionality and questionable reliability. Since there are no published results on analysis by the Rasch model of the 10-item OB-Quest, it is difficult to fully interpret how the Danish translation behaves compared to the original German version (Austria). However, the reliability of the Danish OB-Quest was higher compared to the seven-item German version, which obtained a PSI of 0.51 and a Cronbach’s alpha of 0.38 (7). This might suggest that the revisions of the original scale have been successful. In addition, the thresholds across all ten items were ordered, reflecting that the three score categories worked as intended (27, 33).

In this study, the overall misfit to the Rasch model and how to resolve it was investigated. Misfit was found for two items within the components of challenging and relaxing activities and adaptation of activities according to changed living conditions. Item 1 ‘Do you generally find your activities in your everyday life under-demanding?’ displayed indications of multidimensionality, and item 9 ‘How well can you adapt your activities in your everyday life to changed living conditions, such as a changed state of health’ showed indication of redundancy. Although, items should not be removed based solely on the results of the statistical analysis (27), removal of these two items markedly improved the overall fit statistics and test of unidimensionality. These items were also removed from the original OB-Quest scale in the development phase for the same reasons.
as we found, but without the same improvements in the overall fit statistics and the unidimensionality t-test (7). Based on these findings, the authors of the original OB-Quest scale concluded that the scale was not unidimensional. It is worth noting that Dür et al. (7) applied the t-test approach for unidimensionality testing proposed by Smith (2002), but did not report the proportion of significant test results, which should be below 5% (27, 39). It is therefore difficult to interpret whether or not the original OB-Quest scale was unidimensional. Although unidimensionality is a matter of degree (39), our findings using the t-test approach confirmed acceptable unidimensionality once the fitting solutions had been applied. In the translation process it was discussed if item 9 was too hypothetical for healthy students, but as OB-Quest is generic and has its roots in occupational science, it seemed relevant to capture the relation between OB and changes in health expressed in using adaptation strategies. However, this item has to be rethought in the future.

For OB-Quest/DK, items 1 and 4 displayed uniform DIF in the initial analysis. During the process of solutions for obtaining overall model fit, uniform DIF by gender continued for item 4 ‘How often do you feel overstressed in your everyday life?’ DIF was resolved by splitting the item into two gender-specific items. However, the result might not be reliable and precise due to the small male sample size, as reflected in the SE of 0.42 logit for the male specific item 4 in Table V. Given evidence of differences in perception of stress and coping styles between genders (43), and that stress might be caused by an imbalance in meaningful and valued roles or by being under-occupied or over-occupied (44), it could be suggested that item 4 be modified to encompass aspects directly related to the concept of OB. No DIF was found for age. It was investigated because it was expected that people of varying ages have varying life pressures and challenges. But it might be that cut points other than the age of 23 would have been better, e.g., having young children.
For the OB-Quest/DK, LID was found for item pairs 6 & 7, and 9 & 10. These six items originate from three components in the first version of the OB-Quest (7). The item pairs share aspects from each component of OB and to a great extent contain similar wording, and these facts might explain the findings. LID might be accommodated by grouping the dependent items (35). However, the best solution was found to be the removal of items 1 and 9 showing misfit, which also resulted in decreased LID between items 6 and 7. LID might artificially inflate reliability (35). As the reliability of the OB-Quest/DK is relatively low, the presence of LID between three item pairs out of ten items suggests that the dependent items might benefit from modification, either of the translation per se or the wording of the original items, in agreement with the theoretical framework closely linked to OB. Such a strategy might also contribute to improvement of the misfit items 1 and 9.

During the application of the fitting solutions, the reliability remained relatively low. In addition, some areas of the latent variable were not represented, as identified by the person-item thresholds map. This implies that the OB-Quest/DK scale does not discriminate well between individuals’ different experiences of OB. Since this finding applies both to the original German OB-Quest (7) and the Danish OB-Quest in this study, it might reflect that the scales may require additional items or more response options to increase sensitivity.

In general, the understanding that balance interacts with experiencing a good life has been and is still present in both Western and Eastern thought. However, it is internationally recognised that OB is a complex phenomenon with no universal agreement (10, 44). Various difficulties seem to arise when OB is not clearly defined and it is often interpreted in relation to time spent in general occupational categories, such as work, play, rest and sleep, given that the conceptualization of occupation is contextual; people categorise occupations individually and time may not be the sole deciding factor of importance (45). More focus may be needed on the subjective experiences of
meaningful engagement related to balancing meaningful occupations, occupational roles and purposeful participation in occupations, given that OB seems to provide structure in everyday life along with a sense of purpose, competency and satisfaction (8). For that reason, it may be more applicable and valid to examine self-reported experience of engaging in occupations, irrespective of the main occupational categories developed within OT or occupational science (10, 45) and consider a further development of OB-Quest in light of these reflections.

Methodological considerations

It is a strength that the dual-panel approach was carefully followed. However, it is a limitation that the English version of the OB-quest was translated, given that the English version has not been validated. However, because the German version was taken into account in the translation process and the items were discussed with the developers in the research group, it is assumed that the Danish translation is a valid translation of OB-quest.

One main issue regarding the analysis by the Rasch model is that the low PSI indicates questionable reliability of the fit statistics. Although this study included an adequate sample size of 262 respondents, a larger sample should have been used (41, 42) to allow a subsample to be drawn, to cross-validate the final scale (27). In addition, a larger sample size might have provided a more equal distribution of gender for the DIF analysis and thus a higher degree of precision.

The participants’ responses reflect a current tendency in Danish society. Despite the policy of higher education inspiring new students to choose subjects outside traditional fields of education, it is seen that men’s and women’s educational and career choices continues, with women still in a clear majority in nursing and caring courses and professions (46). Moreover, women continue to assume the main responsibility for caring for children and manage more household chores than do
men, which might hamper OB, and it would have been desirable to assess DIF in relation to these factors (47). This was not possible, however, as the group with children at home was small. Furthermore, we had information only on whether or not the respondent had children at home, but not the age of the children.

It might be that some of the differences between OB-quest and OB-quest/DK relate to the sample of healthy students, given that OB-Quest was generated based on the perspectives of both patients and healthy people. In the future, the psychometric properties of the OB-Quest must be evaluated on more groups to ensure the generic use, e.g., groups of people with reduced functioning.

Clinical implications and future research

The idea behind the OB-Quest is to measure OB and it was developed as a generic self-reported outcome instrument. Furthermore, the summed score gives an overall estimate of the extent to which the individual experiences OB. Since it is an essential aspect of occupational science and OT, it is important to be able to validly assess OB. However, the findings in this study suggest that the OB-Quest is still in its developmental phase and is not quite ready for implementation in clinical practice or research. One main issue seems to be the need for a more elaborated theoretical framework in order to base the instrument on a clear definition of the concept and a conceptual framework consisting of the included components. This may lead to the development of additional relevant items, a broader response scale and further testing in populations with both healthy people and people with disabilities, to aim for a generic tool.

Conclusion

The Danish OB-Quest is in a developmental phase and not ready for implementation in clinical practice or research. One main issue is the need for the further underpinning of the latent variable
(OB) and the development of more items, by applying an appropriate theoretical framework, to develop coherent measurements of the self-reported experience of OB.
References


46. Myrup MS/KINFO [The Danish Center for Research on Women and Gender]. Fortsat skæv kønsfordeling [Continued skewed gender distribution] [Internet]. KVINFO Køn Viden Information Forskning [The Danish Center for Research on Women and Gender]; 2019 [cited 20.08.06]. Available from: https://kvinfo.dk/fortsat-skæv-kønsfordeling-paa-videregaaende-studier/

47. Danmarks Statistik [Statistics Denmark]. Ligestilling i Danmark: Ligestillingswebsite [Gender equality in Denmark: Gender equality website] [Internet]. [cited 20.08.06] Available from: https://www.dst.dk/da/Statistik/emner/levevilkaar/ligestilling/ligestillingswebsite
List of figures and legends

Figure 1: Item characteristic curves (ICC) of misfitting items of OB-Quest/DK

Legends figure 1

ICC plot for two items. Based on the sample size, persons (respondents) are divided into four groups (class intervals with at least 50 persons in each). The curved line represents the expected scores for the item, and the dots represent the observed scores for the class intervals at the different levels of the measured variable (OB). A high score is indicative of a low degree of OB. A) The ICC plot for item 1 with a high positive fit residual of 2.6, which is significant. The observed scores form a flatter curve than the expected scores, which indicates that this item is under-discriminating and might reflect multidimensionality. B) The ICC plot for item 9 with a medium high negative fit residual of -1.84. The observed scores form a steeper curve than the expected scores, which indicates that this item is over-discriminating and might reflect potential redundancy or dependency within the item set.

Figure 2: Uniform differential item function (DIF) for item 1 (Activities under-demanding) and item 4 (Feel overstressed)

Legends figure 2

A) ICC plot for item 1 (Activities under-demanding) and the observed responses from occupational therapy students (blue) and other health professional students (red), which shows evidence of uniform DIF with occupational therapy students scoring more activities under-demanding than other students, given the same location on the latent variable. B) ICC plot for item 4 (Feel overstressed) and the observed responses from females (blue) and males (red), which shows evidence of uniform DIF, with females scoring more feeling of overstress than males,
given the same location on the latent variable. Due to the small sample size for males, the number of ability groups was reduced from four to two class intervals.

Figure 3: Person-Item threshold distribution of the OB-Quest/DK data

Legends figure 3

The graphs display the person-item threshold distribution map for: A) the OB-Quest/DK with eight items, before resolving DIF by gender for item 4, and B) the OB-Quest/DK with nine items, after resolving DIF by gender for item 4. The x-axes display location of item thresholds (lower half) and location of OB reported on the OB-Quest/DK by respondents (upper half). High score is indicative of low degrees of OB and low scores indicate high degrees of OB. The y-axes display the frequencies of item thresholds (lower half) and respondents (upper half). For both graphs, the item thresholds spread over six logits with no evidence of floor or ceiling effects. Some item-thresholds are in the same place, which indicates that they are duplicating the ability to discriminate at that level of difficulty. Some areas along the logit scale are not represented by item-thresholds.
### Table I Items in OB-Quest

1. Do you generally find your activities in your everyday life under-demanding?
2. Do you generally find your activities in your everyday life over-demanding?
3. Do you generally receive enough appreciation for activities in your everyday life?
4. How often do you feel overstressed in your everyday life?
5. How much are your activities in your everyday life affected by your health?
6. Do you get enough rest?
7. Do you get enough sleep?
8. Do you have sufficient variety of different activities that you do? For example, do you do a mixture of physical activities and more sedentary ones (where you are sitting down or staying still)? Or a mixture of creative activities and activities that are more routine for you?
9. How well can you adapt your activities in your everyday life to changed living conditions, such as a changed state of health?
10. How well can you adapt your activities in your everyday life to changed living conditions, such as a change of your professional life or employment status?
Table II Demographics of respondents (N=262)

<table>
<thead>
<tr>
<th>Subgroups for the analysis of DIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender: female /male (N) 237 / 25</td>
</tr>
<tr>
<td>Age Mean (SD) 26.10 (7.05) years</td>
</tr>
<tr>
<td>Age min / max 19 / 53 years</td>
</tr>
<tr>
<td>Age median 23 years</td>
</tr>
<tr>
<td>Bachelor (BA) programme (N)</td>
</tr>
<tr>
<td>Occupational therapy (OT) 132</td>
</tr>
<tr>
<td>Multi media / IT 15</td>
</tr>
<tr>
<td>Nursing 115</td>
</tr>
<tr>
<td>Time in education (N)</td>
</tr>
<tr>
<td>Semester 1 0</td>
</tr>
<tr>
<td>Semester 2 71</td>
</tr>
<tr>
<td>Semester 3 69</td>
</tr>
<tr>
<td>Semester 4 15</td>
</tr>
<tr>
<td>Semester 5 90</td>
</tr>
<tr>
<td>Semester 6 16</td>
</tr>
<tr>
<td>Semester 7 1</td>
</tr>
<tr>
<td>Work in addition to study</td>
</tr>
<tr>
<td>Yes /no 188/74</td>
</tr>
<tr>
<td>Time used for daily transportation</td>
</tr>
<tr>
<td>Time mean (SD) 32.82 (29.27) minutes</td>
</tr>
<tr>
<td>Time min/max 2 / 225 minutes</td>
</tr>
<tr>
<td>Time median 20 minutes</td>
</tr>
<tr>
<td>Children at home</td>
</tr>
<tr>
<td>Yes /no 49/213</td>
</tr>
<tr>
<td>Civil status</td>
</tr>
<tr>
<td>Married 37</td>
</tr>
<tr>
<td>In a relationship 142</td>
</tr>
<tr>
<td>Single 80</td>
</tr>
<tr>
<td>Divorced 3</td>
</tr>
</tbody>
</table>
Table III Analysis by the Rasch model – overall fit statistics for the OB-Quest/DK

<table>
<thead>
<tr>
<th>Scale Analysis</th>
<th>No. of Items</th>
<th>Fit residuals Mean (SD)</th>
<th>Item-trait interaction</th>
<th>Reliability</th>
<th>Unidimensionality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Item</td>
<td>Person</td>
<td>$\chi^2$ (df)</td>
<td>P</td>
</tr>
<tr>
<td>Analysis 1: initial analysis</td>
<td>10</td>
<td>0.08 (1.27)</td>
<td>-0.19 (1.06)</td>
<td>74.12 (30)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Analysis 2: remove misfit item 1</td>
<td>9</td>
<td>-0.01 (1.07)</td>
<td>-0.24 (1.11)</td>
<td>38.14 (27)</td>
<td>0.076</td>
</tr>
<tr>
<td>Analysis 3: remove misfit item 9</td>
<td>8</td>
<td>0.09 (0.87)</td>
<td>-0.25 (1.09)</td>
<td>28.07 (24)</td>
<td>0.257</td>
</tr>
<tr>
<td>Analysis 4: split item 4 for DIF by gender</td>
<td>9</td>
<td>0.14 (0.91)</td>
<td>-0.24 (1.09)</td>
<td>29.38 (27)</td>
<td>0.347</td>
</tr>
</tbody>
</table>

**Optimal fit**

|                      | 0 (< 1.00) | 0 (< 1.00) | > 0.05 | > 0.70 | <5% or lower CI <5% |

Abbreviations: OB-Quest/DK, Danish version of the Occupational Balance Questionnaire; (SD), Standard deviation; ($\chi^2$), Chi-square; (df), degrees of freedom; PSI, Person Separation Index; $\alpha$, Cronbach’s alpha; CI, Confidence interval; NA, Not applicable.
### Table IV. Analysis by the Rasch model – item level fit statistics for the OB-Quest/DK

<table>
<thead>
<tr>
<th>Item and label (abbreviated)</th>
<th>Loc</th>
<th>SE</th>
<th>FR</th>
<th>$\chi^2$ (df)</th>
<th>$P^*$</th>
<th>F (df)</th>
<th>$P^*$</th>
<th>Threshold</th>
<th>Uniform DIF</th>
<th>Non-Uniform DIF</th>
<th>LID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Activities under-demanding</td>
<td>0.34</td>
<td>0.13</td>
<td>2.6</td>
<td>18.04 (3)</td>
<td>&lt;0.001</td>
<td>5.55 (3,257)</td>
<td>0.001</td>
<td>Ordered</td>
<td>Education</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2. Activities over-demanding</td>
<td>0.37</td>
<td>0.14</td>
<td>0.8</td>
<td>4.55 (3)</td>
<td>0.208</td>
<td>1.55 (3,257)</td>
<td>0.203</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>3. Receive appreciation</td>
<td>-0.51</td>
<td>0.11</td>
<td>0.7</td>
<td>2.06 (3)</td>
<td>0.560</td>
<td>0.68 (3,257)</td>
<td>0.564</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4. Feel overstressed</td>
<td>-1.29</td>
<td>0.12</td>
<td>-0.8</td>
<td>5.43 (3)</td>
<td>0.143</td>
<td>2.22 (3,257)</td>
<td>0.086</td>
<td>Ordered</td>
<td>Gender</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>5. Activities affected by health</td>
<td>-0.45</td>
<td>0.11</td>
<td>-0.1</td>
<td>1.59 (3)</td>
<td>0.662</td>
<td>0.55 (3,257)</td>
<td>0.651</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>6. Get enough rest</td>
<td>0.04</td>
<td>0.11</td>
<td>0.6</td>
<td>3.09 (3)</td>
<td>0.378</td>
<td>0.85 (3,257)</td>
<td>0.466</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>7. Get enough sleep</td>
<td>0.37</td>
<td>0.12</td>
<td>-0.1</td>
<td>0.58 (3)</td>
<td>0.901</td>
<td>0.10 (3,257)</td>
<td>0.961</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>0.26</td>
</tr>
<tr>
<td>8. Variety of activities</td>
<td>0.25</td>
<td>0.12</td>
<td>0.5</td>
<td>0.84 (3)</td>
<td>0.840</td>
<td>0.24 (3,257)</td>
<td>0.871</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>9. Adapt activity-health</td>
<td>0.81</td>
<td>0.13</td>
<td>-1.8</td>
<td>22.73 (3)</td>
<td>&lt;0.001</td>
<td>11.16 (3,257)</td>
<td>&lt;0.001</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>10. Adapt activity-living condition</td>
<td>0.05</td>
<td>0.13</td>
<td>-1.5</td>
<td>15.19 (3)</td>
<td>0.002</td>
<td>7.18 (3,257)</td>
<td>&lt;0.001</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Abbreviations: OB-Quest/DK, Danish version of the Occupational Balance Questionnaire; Loc, location; SE, Standard error; FR, fit residual; $\chi^2$, Chi-square; df, degrees of freedom; DIF, differential item function; LID, local item independence.

Bold highlighted with grey indicates misfit or violation of the Rasch model.

*Bonferroni adjusted $p = 0.005$ for ten items.
<table>
<thead>
<tr>
<th>Item and label (abbreviated)</th>
<th>Loc</th>
<th>SE</th>
<th>FR</th>
<th>$\chi^2$ (df)</th>
<th>$P^*$</th>
<th>F (df)</th>
<th>$P^*$</th>
<th>Threshold</th>
<th>Uniform DIF</th>
<th>Non-Uniform DIF</th>
<th>LID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>With item items 1 and 9 removed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Activities over-dim Demanding</td>
<td>0.52</td>
<td>0.14</td>
<td>1.0</td>
<td>2.55 (3)</td>
<td>0.466</td>
<td>0.87 (3,257)</td>
<td>0.458</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>3. Receive appreciation</td>
<td>-0.38</td>
<td>0.11</td>
<td>0.9</td>
<td>3.40 (3)</td>
<td>0.333</td>
<td>1.23 (3,257)</td>
<td>0.299</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4. Feel overstressed</td>
<td>-1.16</td>
<td>0.12</td>
<td>-0.8</td>
<td>4.90 (3)</td>
<td>0.180</td>
<td>2.13 (3,257)</td>
<td>0.096</td>
<td>Ordered</td>
<td>Gender</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>5. Activities affected by health</td>
<td>-0.31</td>
<td>0.11</td>
<td>0.2</td>
<td>1.01 (3)</td>
<td>0.799</td>
<td>0.35 (3,257)</td>
<td>0.789</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>6. Get enough rest</td>
<td>0.22</td>
<td>0.11</td>
<td>0.02</td>
<td>2.19 (3)</td>
<td>0.535</td>
<td>0.82 (3,257)</td>
<td>0.484</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>7. Get enough sleep</td>
<td>0.54</td>
<td>0.12</td>
<td>-0.9</td>
<td>2.33 (3)</td>
<td>0.507</td>
<td>0.95 (3,257)</td>
<td>0.419</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>8. Variety of activities</td>
<td>0.39</td>
<td>0.12</td>
<td>1.2</td>
<td>3.40 (3)</td>
<td>0.334</td>
<td>1.06 (3,257)</td>
<td>0.366</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>10. Adapt activity-living condition</td>
<td>0.18</td>
<td>0.13</td>
<td>-0.8</td>
<td>8.30 (3)</td>
<td>0.040</td>
<td>3.55 (3,257)</td>
<td>0.015</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>With item 1 and 9 removed and item 4 split by gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Activities over-dim Demanding</td>
<td>0.49</td>
<td>0.14</td>
<td>1.1</td>
<td>3.57 (3)</td>
<td>0.311</td>
<td>1.24 (3,257)</td>
<td>0.296</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>3. Receive appreciation</td>
<td>-0.40</td>
<td>0.12</td>
<td>1.0</td>
<td>3.85 (3)</td>
<td>0.278</td>
<td>1.40 (3,257)</td>
<td>0.245</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4. Feel overstressed-Female</td>
<td>-1.35</td>
<td>0.13</td>
<td>-1.1</td>
<td>6.03 (3)</td>
<td>0.110</td>
<td>2.75 (3,257)</td>
<td>0.043</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4. Feel overstressed-Male</td>
<td>0.32</td>
<td>0.42</td>
<td>0.6</td>
<td>0.49 (3)</td>
<td>0.921</td>
<td>0.14 (3,20)</td>
<td>0.934</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>5. Activities affected by health</td>
<td>-0.34</td>
<td>0.11</td>
<td>0.2</td>
<td>0.73 (3)</td>
<td>0.867</td>
<td>0.19 (3,257)</td>
<td>0.904</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>6. Get enough rest</td>
<td>0.20</td>
<td>0.11</td>
<td>0.0</td>
<td>3.62 (3)</td>
<td>0.305</td>
<td>1.40 (3,257)</td>
<td>0.242</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>7. Get enough sleep</td>
<td>0.52</td>
<td>0.12</td>
<td>-1.0</td>
<td>2.22 (3)</td>
<td>0.529</td>
<td>0.93 (3,257)</td>
<td>0.427</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>8. Variety of activities</td>
<td>0.38</td>
<td>0.12</td>
<td>1.3</td>
<td>2.91 (3)</td>
<td>0.406</td>
<td>0.88 (3,257)</td>
<td>0.450</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>10. Adapt activity-living condition</td>
<td>0.17</td>
<td>0.13</td>
<td>-0.8</td>
<td>5.97 (3)</td>
<td>0.113</td>
<td>2.50 (3,257)</td>
<td>0.060</td>
<td>Ordered</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Abbreviations: OB-Quest/DK, Danish version of the Occupational Balance Questionnaire; DIF, differential item function; Loc, location; SE, Standard error; FR, fit residual; $\chi^2$, Chi-square; df, degrees of freedom; LID, local item independence.

Bold highlighted with grey indicates misfit or violation of the Rasch model.

*Bonferroni adjusted $p = 0.006$ for eight items or nine items
Figure 1. Item characteristic curves (ICC) of mis-fitting items of OB-Quest/DK

ICC plot for two items. Based on the sample size, persons (respondents) are divided into four groups (class intervals with at least 50 persons in each). The curved line represents the expected scores for the item, and the dots represent the observed scores for the class intervals at the different levels of the measured variable (OB). A high score is indicative of a low degree of OB. A) The ICC plot for item 1 with a high positive fit residual of 2.6, which are significant. The observed scores form a flatter curve than the expected scores, which indicates that this item is under-discriminating and might reflect multidimensionality. B) The ICC plot for item 9 with a medium high negative fit residual of -1.84. The observed scores form a steeper curve than the expected scores, which indicates that this item is over-discriminating and might reflect potential redundancy or dependency within the item set.
Figure 2. Uniform differential item function (DIF) for item 1 (Activities under-demanding) and item 4 (Feel overstressed).

A) ICC plot for item 1 (Activities under-demanding) and the observed responses from occupational therapy students (blue) and other health professional students (red), which shows evidence of uniform DIF with occupational therapy students scoring more activities under-demanding than other students given the same location on the latent variable. B) ICC plot for item 4 (Feel overstressed) and the observed responses from females (blue) and males (red), which shows evidence of uniform DIF with females scoring more feeling of overstress than males given the same location on the latent variable. Due to the small sample size for males, the number of ability groups are reduced from four to two class intervals.
Figure 3. Person-Item threshold distribution of the OB-Quest/DK data.

The graphs display the person-item threshold distribution map for A) the OB-Quest/DK with eight items before resolving DIF by gender for item 4, and B) the OB-Quest/DK with nine items after resolving DIF by gender for item 4. The x-axes display location of item thresholds (lower half) and location of OB reported on the OB-Quest/DK by respondents (upper half). High score is indicative of low degrees of OB and low scores indicates high degrees of OB. The y-axes display the frequencies of item thresholds (lower half) and respondents (upper half). For both graphs, the item thresholds spread over 6 logits with no evidence of floor or ceiling effects. Some item-thresholds are in the same place, which indicates that they are duplicating the ability to discriminate at that level of difficulty. Some areas along the logit scale are not represented by item-thresholds.