Construct validity of a revised version of the Occupational Balance Questionnaire

Carita Håkansson, Petra Wagman & Peter Hagell

To cite this article: Carita Håkansson, Petra Wagman & Peter Hagell (2019): Construct validity of a revised version of the Occupational Balance Questionnaire, Scandinavian Journal of Occupational Therapy, DOI: 10.1080/11038128.2019.1660801

To link to this article: https://doi.org/10.1080/11038128.2019.1660801

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 14 Sep 2019.

Submit your article to this journal

Article views: 143

View related articles

View Crossmark data
Construct validity of a revised version of the Occupational Balance Questionnaire

Carita Håkansson, Petra Wagman and Peter Hagell

Division of Occupational and Environmental Medicine, Lund University, Lund, Sweden; Department of Rehabilitation, School of Health and Welfare, Jönköping University, Jönköping, Sweden; Department of Nursing and Integrated Health Sciences, Kristianstad University, Kristianstad, Sweden

ABSTRACT

Background: Occupational balance is fundamental to occupational therapy and occupational science. Therefore, the Occupational Balance Questionnaire (OBQ) was developed and has previously been found psychometrically valid according to classical test theory.

Aim: To investigate the internal construct validity of the OBQ using Rasch measurement theory.

Material and methods: Data from two general population samples were used to investigate the psychometric properties of the OBQ according to Rasch measurement theory.

Results: The analyses identified problems with the current response scale and multidimensionality of two items. As a result, a revised version, the OBQ11, was suggested and exhibited response categories that worked properly, good reliability (0.92), model fit and measurement invariance across age and gender groups. The hierarchical item ordering was in agreement with previous research.

Conclusion: The new OBQ11 satisfies the measurement criteria defined by the Rasch model. However, further studies of additional samples are needed to validate its generic properties.

Significance: The purpose of the OBQ11 is to measure occupational balance of individuals or groups, and to identify aspects of occupational balance in need of improvement. Our observations suggest that the OBQ11 is a valid and promising complement to other instruments.

ARTICLE HISTORY
Received 3 May 2018
Revised 13 August 2019
Accepted 23 August 2019

KEYWORDS
Measurement; occupational science; occupational therapy; psychometrics

Introduction

Occupational balance has been fundamental to occupational therapy theory and practice ever since the birth of the profession [1]. Occupational balance is not focusing on balancing different occupations into equal proportions or equilibrium. Rather it is a complex concept and it is conceptualized and measured in different ways, focusing on various aspects such as need satisfaction [2–3], levels of meaningfulness [4], and types of occupation [5], as described below.

Matuska [2] developed the Life Balance Inventory (LBI) conceptualizing life balance, based on the life balance model [3] and defines it as ‘a satisfying pattern of daily activity that is healthful, meaningful and sustainable to an individual within the context of his or her current situation’ (p. 11). LBI consists of 53 predefined activities and people rate how satisfied they are with how much time they spent doing the activity in the last month. Furthermore, balance is assessed as congruence and equivalence, where congruence means a match between actual and desired activity configuration, and equivalence means approximately equal apportioning of time use across the four needs dimensions. Eakman [4] developed the Meaningful Activity Wants and Needs Assessment (MAWNA), and he defined life balance as high levels of meaningful activity concurrently with low needs for additional meaningful activity experiences [4]. MAWNA assesses the balance between actual and desired engagement in meaningful occupations. Dür et al. [5] developed the Occupational Balance Questionnaire (OB-Quest) among people with autoimmune diseases and healthy people. They found that occupational balance includes the following components: balance between different types of occupations, between abilities and resources, and between one’s own and others’ needs. Based on these components 10 items were developed.

CONTACT Carita Håkansson carita.hakansson@med.lu.se Division of Occupational and Environmental Medicine, Lund University, Medicon Village, Lund SE 223 81, Sweden

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.
However, another important aspect of occupational balance, not yet addressed in a measurement is ‘the experience of having the right amount of occupations and the right variation between occupations in the occupational pattern’ [6, p. 326] (regardless of what the occupations are). Therefore, we developed the Occupational Balance Questionnaire (OBQ) [7]. It intends to be a generic instrument to evaluate occupational balance of individuals and groups. In occupational therapy a valid and reliable instrument to measure this conceptualization of occupational balance is needed. The conceptual framework for the OBQ is based on results from previous research on the experience of occupational balance in different groups [8–10], and a concept analysis of occupational balance [6]. Types of indicators that are useful for eliciting occupational balance were developed based on these studies (potential item format). The 13 OBQ items (Table 1) have six ordered response categories with verbal descriptions and the participant rates his/her level of agreement with different aspects of amount of and variation between occupations in the occupational pattern (0 = completely disagree, 1 = disagree a lot, 2 = tend to disagree, 3 = tend to agree, 4 = agree a lot, and 5 = completely agree). Item scores are summed into a total score that can range from 0 to 65. A higher score means a higher occupational balance. The OBQ focuses on the variation in the occupational pattern (between different types of occupations, between doing things for oneself and for others, between doing things alone and together with others), the amount of each occupation and the total amount of occupations (amount of, time for, number of), and the perceived meaningfulness in the occupations. All these aspects are considered necessary for the experience of occupational balance.

The OBQ was piloted using classical test theory (CTT) to evaluate its psychometric properties [7]. Results showed good total score internal consistency (Cronbach alpha, 0.94) and test-retest reliability (Spearman’s Rho, 0.93) without any floor or ceiling effects. Test-retest reliability of individual items varied between 0.61–0.83.

However, CTT approaches to assessing the measurement properties of scales have limitations [11–13]. For example, it rests on assumptions that cannot be readily tested and is largely based on correlation analyses, which has limited relevance to the concept of measurement as understood in physical sciences as well as in common language (e.g. in terms of common ruler) [12,13]. Additionally, CTT does not readily allow for assessment of fundamental aspects of the measurement process, such as whether response categories work as expected and whether the instrument works in the same manner across different subsets of respondents. To this end, the Rasch measurement model [14] provides a powerful alternative by determining the extent to which observed data confirm with the model [11–15]. The Rasch model is a general unidimensional measurement model for latent variables that is based on the same principles as measurements in physical sciences; the model states that the probability of a certain item response depends on the level of the measured construct represented by the item and the person, respectively [16,17]. As such, it provides a detailed account for the extent to which item responses are productive in producing objective measures at an interval level, as well as a powerful diagnostic tool regarding any anomalies that contradict objective measurement (e.g. whether response categories work as intended and if items are working the same way across respondent subgroups). Furthermore, the Rasch model determines empirically how items (and persons) are ordered on a latent continuum from less to more of the target variable, which allows for an investigation of the coherence between the instrument and its underpinning theory [18,19]. As such, Rasch measurement theory (RMT) provides a valuable and theoretically as well as clinically appropriate tool in assessing the

<table>
<thead>
<tr>
<th>Item number</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Having just enough to do during a regular week</td>
</tr>
<tr>
<td>2</td>
<td>Balance between doing things for others and for oneself</td>
</tr>
<tr>
<td>3</td>
<td>Perceiving one’s occupations as meaningful</td>
</tr>
<tr>
<td>4</td>
<td>Time for doing things wanted</td>
</tr>
<tr>
<td>5</td>
<td>Balance between work, home, family, leisure, rest, and sleep</td>
</tr>
<tr>
<td>6</td>
<td>Balance between doing things alone and with others</td>
</tr>
<tr>
<td>7</td>
<td>Have sufficient time for doing obligatory occupations</td>
</tr>
<tr>
<td>8</td>
<td>Balance between physical, social, mental, and restful occupations</td>
</tr>
<tr>
<td>9</td>
<td>Satisfaction with how time is spent in everyday life</td>
</tr>
<tr>
<td>10</td>
<td>Satisfaction with the number of activities during a regular week</td>
</tr>
<tr>
<td>11</td>
<td>Balance between obligatory and voluntary occupations</td>
</tr>
<tr>
<td>12</td>
<td>Balance between energy-giving and energy-taking activities</td>
</tr>
<tr>
<td>13</td>
<td>Satisfaction with time spent in rest, recovery, and sleep</td>
</tr>
</tbody>
</table>
measurement properties of rating scale based instruments [11–15,18].

**Aim**

The aim of the present study was to investigate the internal construct validity of the OBQ in a general population using Rasch measurement theory. Specifically, item fit, reliability, response categories, differential item functioning and logical hierarchical ordering of the items were assessed.

**Material and methods**

**Participants**

As this article reports a process, data collection was conducted in two phases and included two samples (Samples 1 and 2) from general populations in Southern Sweden. They were selected through purposive sampling. The authors and their students used their network to identify presumptive participants, and they received an information letter that invited them to the study, detailing the aim and study design. It also explained that it was voluntary to take part in the study and that the participants could withdraw from the study at any time without giving any explanation. Finally, the letter explained that the participants answered the OBQ anonymously and that no unauthorized person would have access to the data. An addressed return enveloped was also added. The inclusion criteria were persons 20 years old and older (as they are no longer in school) and the ability to understand Swedish. A heterogeneous sample regarding age and gender was the goal.

**Data analyses**

In this study, the OBQ was tested according to RMT (12, 14, 15). Specifically, reliability, Rasch model fit, rating scale response category functioning, uniform and non-uniform Differential Item Functioning (DIF) by gender and age (younger vs. older persons, as defined by the median), targeting, and the logic of the hierarchical item ordering were addressed. Reliability was estimated using the Pearson Separation Index (PSI), which is conceptually analogous to coefficient alpha [20], and should preferably be at least 0.8 [12]. Fit between empirical data and the unidimensional Rasch measurement model was examined from several perspectives. Individual item fit residuals represent discrepancies between empirical responses and model expectations and should range between about −2.5 and 2.5, with the ideal being 0 [21]. Overall, the mean standardized item and person fit residuals should be close to 0 with a standard deviation close to 1. Chi-square statistics test whether the difference between the observed and expected values for each item is statistically significant or not. In general, non-significant Chi-square statistic supports fit to the model. In addition, accordance between item response data and model expectations was examined graphically by inspection of item characteristic curves (ICC). Taken together, these analyses provide evidence of potential anomalies between any particular item which may over- or under discriminate relative to the summary discrimination of all items [22]. Anomalies may arise due to measurement problems such as local response dependence, multidimensionality or poor item quality, which may threaten validity [19]. Large negative standardized fit residuals suggest over-discriminating items and local response dependence, whereas large positive residual values indicate under-discriminating items and are typically due to local trait dependence (i.e. multidimensionality) [12]. To examine local dependency in greater detail, the relative correlations between standardized item residuals were examined. In general, these correlations should not exceed 0.2–0.25 [23,24].

The OBQ has response categories ranging from low agreement to high agreement with the statements in the questionnaire. Whether the OBQ response categories work as assumed was assessed by examining response category thresholds [22]. A threshold is the point where the probability of a response is the same between two adjacent response categories. In the event of disordered thresholds, response categories do not work as intended.

DIF is an additional aspect of fit that concerns whether items work the same way for different subgroups, for example different age groups or men and women. Age was dichotomized into older (50–89 years old) vs. younger (20–49 years old) according to the median. Here we addressed uniform (subgroup main effect) and non-uniformed (subgroup × location interaction effect) DIF by 2-way ANOVAs of residuals [25]. Items exhibiting DIF were adjusted by splitting the item into two subgroup-specific items followed by examination of the impact of DIF on the total score.

Targeting refers to the correspondence between item and person locations, where person locations along the measurement continuum preferably should be mirrored by item locations, and vice versa, in order to maximize precision in estimated person and
item measures as well as the conditions for scale evaluation. Detecting differences between people is an important function of measurement, which is linked to measurement precision and, therefore, also to targeting and reliability.

According to RMT, people with low occupational balance are expected to have low probabilities to endorse items expressing high levels of occupational balance. As this is the basis of locating items (and persons) on the latent logit continuum, it provides the foundations for examining the nature of how the construct varies as one moves from lower to higher levels of occupational balance, and the extent to which this is in accord with clinical and theoretical experiences. The logic of the hierarchical ordering of item locations was, therefore, considered as means to assess internal content and construct validity.

Data were analyzed according to the unrestricted polytomous (‘partial credit’) Rasch model using RUMM2030 [26], and the program divided the samples into four class intervals (subgroups of people with similar levels of occupational balance according to total OBQ scores). Bonferroni adjustments for multiple null hypothesis testing were applied, with the alpha level of significance set at 0.05 [27,28].

**Results**

Sample 1, (who responded to the original 13-item version of the OBQ), consisted of 168 persons aged 21–64. The mean (SD) age was 53 (12) and 68 per cent were women. The reliability of the OBQ was good, and the overall fit to the model was reasonable (Table 2). However, the response category thresholds were disordered in four items (items 2, 4, 6 and 7). A similar pattern (albeit marginally ordered) was also observed for a majority of the remaining items. This suggests that respondents had difficulties distinguishing between response categories, threatens both the user-friendliness of the scale and the interpretability of responses, and was likely related to too many response categories. Other reasons may also include multidimensionality [19], which appeared to be the case for three items, or local response dependence [23,29], which was indicated for four items (Table 3). Based on these observations, we explored the effects of step-wise collapsing response categories. First, we collapsed the first and second response categories and after that also the third and fourth response categories. Following this, the rating scale worked as intended (data not shown). Thus, the response categories were reduced from six to four (0 = completely disagree, 1 = tend to disagree, 2 = tend to agree, and 3 = completely agree).

New data were then collected with this revised version of the OBQ (Sample 2). Sample 2 consisted of 207 persons aged 20–89. The mean (SD) age was 50 (13) and 64 per cent were women. As suggested by our observations with the experimentally collapsed response categories in sample 1, there were no disordered thresholds in sample 2 (Figure 1). The revised OBQ with four response categories showed good reliability and improved overall fit (Table 2). Table 3 shows individual item fit statistics for the OBQ with 13 items following reduction to four response categories. Three items showed signs of misfit. In two of these (items 3 and 6) fit residuals outside the range (4.034 and 2.902, respectively) suggest multidimensionality. These items were, therefore, reviewed relative to the definition of occupational balance, i.e. ‘the experience of having the right amount of occupations and the right variation between occupations in the

**Table 2.** Overall Rasch model fit statistics and reliability.

<table>
<thead>
<tr>
<th></th>
<th>Sample 1. 13 items, 6 response categories</th>
<th>Sample 2. 13 items, 4 response categories</th>
<th>Sample 2. 11 items, 4 response categories (OBQ11)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item fit residuals</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>−0.148</td>
<td>−0.039</td>
<td>−0.025</td>
</tr>
<tr>
<td>SD</td>
<td>3.254</td>
<td>1.793</td>
<td>1.149</td>
</tr>
<tr>
<td><strong>Person fit residuals</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>−0.641</td>
<td>−0.310</td>
<td>−0.377</td>
</tr>
<tr>
<td>SD</td>
<td>1.795</td>
<td>1.345</td>
<td>1.488</td>
</tr>
<tr>
<td><strong>Total item-trait interaction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total item Chi2 (df)</td>
<td>121.16 (39)</td>
<td>86.48 (39)</td>
<td>18.21 (22)</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.693</td>
</tr>
<tr>
<td>Reliability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person separation index</td>
<td>0.93</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>Targeting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person location, mean (SD)</td>
<td>1.331 (1.540)</td>
<td>0.480 (1.750)</td>
<td>0.300 (1.829)</td>
</tr>
</tbody>
</table>

*aShould be close to 0.*

*bShould be close to 1.*

*cShould be >0.05 to suggest model fit at all levels of balance and over all items.*

*dPerson Separation Index, should be > 0.8, relative to the mean item logit location (i.e. zero).*

C. HÅKANSSON ET AL.
### Table 3. Rasch item and fit statistics.

<table>
<thead>
<tr>
<th>Item</th>
<th>Location order</th>
<th>Item location (SE)</th>
<th>Fit residual *</th>
<th>Chi-sq * (p-value)</th>
<th>Item</th>
<th>Location order</th>
<th>Item location (SE)</th>
<th>Fit residual *</th>
<th>Chi-sq * (p-value)</th>
<th>Item</th>
<th>Location order</th>
<th>Item location (SE)</th>
<th>Fit residual *</th>
<th>Chi-sq * (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>0.178 (0.052)</td>
<td>-0.672</td>
<td>0.761 (0.859)</td>
<td>10</td>
<td>0.516 (0.113)</td>
<td>0.320</td>
<td>2.573 (0.276)</td>
<td>9</td>
<td>0.304 (0.119)</td>
<td>1.080</td>
<td>1.546 (0.462)</td>
<td>6</td>
<td>0.672 (0.646)</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>-0.234 (0.055)</td>
<td>3.552</td>
<td>19.358 (&lt;0.001)</td>
<td>4</td>
<td>-0.423 (0.054)</td>
<td>0.918</td>
<td>4.043 (0.307)</td>
<td>1</td>
<td>0.297 (0.119)</td>
<td>0.874</td>
<td>1.546 (0.462)</td>
<td>12</td>
<td>0.673 (0.696)</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.234 (0.055)</td>
<td>3.552</td>
<td>19.358 (&lt;0.001)</td>
<td>13</td>
<td>0.774 (0.130)</td>
<td>0.768</td>
<td>3.453 (0.178)</td>
<td>13</td>
<td>0.378 (0.129)</td>
<td>-0.168</td>
<td>2.761 (0.251)</td>
<td>12</td>
<td>0.724 (0.696)</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.690 (0.056)</td>
<td>1.866</td>
<td>3.074 (0.337)</td>
<td>14</td>
<td>0.577 (0.123)</td>
<td>-1.659</td>
<td>3.682 (0.159)</td>
<td>2</td>
<td>1.022 (0.135)</td>
<td>1.905</td>
<td>2.498 (0.287)</td>
<td>13</td>
<td>0.874 (0.287)</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>0.224 (0.052)</td>
<td>-1.274</td>
<td>3.375 (0.337)</td>
<td>15</td>
<td>0.305 (0.116)</td>
<td>2.263</td>
<td>11.962 (0.003)</td>
<td>11</td>
<td>0.378 (0.129)</td>
<td>-0.168</td>
<td>2.761 (0.251)</td>
<td>13</td>
<td>0.904 (0.113)</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>-0.098 (0.053)</td>
<td>1.566</td>
<td>3.036 (0.366)</td>
<td>16</td>
<td>0.111 (0.121)</td>
<td>-0.521</td>
<td>0.495 (0.781)</td>
<td>7</td>
<td>0.332 (0.126)</td>
<td>-0.642</td>
<td>0.912 (0.634)</td>
<td>13</td>
<td>0.912 (0.634)</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>-0.074 (0.051)</td>
<td>3.866</td>
<td>17.711 (&lt;0.001)</td>
<td>17</td>
<td>-0.035 (0.053)</td>
<td>-0.130</td>
<td>0.035 (0.786)</td>
<td>4</td>
<td>0.111 (0.127)</td>
<td>-0.432</td>
<td>0.483 (0.786)</td>
<td>14</td>
<td>0.904 (0.113)</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0.157 (0.055)</td>
<td>-3.141</td>
<td>6.767 (0.080)</td>
<td>18</td>
<td>-0.427 (0.125)</td>
<td>-0.884</td>
<td>2.207 (0.311)</td>
<td>5</td>
<td>0.199 (0.130)</td>
<td>-0.020</td>
<td>0.930 (0.628)</td>
<td>14</td>
<td>0.930 (0.628)</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>0.017 (0.054)</td>
<td>-4.683</td>
<td>18.159 (&lt;0.001)</td>
<td>19</td>
<td>0.502 (0.122)</td>
<td>-2.133</td>
<td>11.962 (0.003)</td>
<td>7</td>
<td>0.332 (0.126)</td>
<td>-0.642</td>
<td>0.912 (0.634)</td>
<td>13</td>
<td>0.912 (0.634)</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>0.243 (0.049)</td>
<td>0.696</td>
<td>2.479 (0.479)</td>
<td>20</td>
<td>-0.375 (0.126)</td>
<td>-0.035</td>
<td>2.265 (0.113)</td>
<td>8</td>
<td>0.300 (0.127)</td>
<td>-1.783</td>
<td>4.352 (0.113)</td>
<td>13</td>
<td>0.904 (0.113)</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>0.200 (0.053)</td>
<td>-5.646</td>
<td>18.705 (&lt;0.001)</td>
<td>21</td>
<td>0.343 (0.121)</td>
<td>-0.113</td>
<td>3.472 (0.176)</td>
<td>10</td>
<td>0.109 (0.127)</td>
<td>-0.064</td>
<td>2.209 (0.331)</td>
<td>13</td>
<td>0.912 (0.634)</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>0.439 (0.053)</td>
<td>-3.105</td>
<td>11.532 (0.009)</td>
<td>22</td>
<td>0.536 (0.121)</td>
<td>-1.706</td>
<td>5.418 (0.067)</td>
<td>12</td>
<td>0.332 (0.126)</td>
<td>-0.642</td>
<td>0.912 (0.634)</td>
<td>13</td>
<td>0.912 (0.634)</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>0.063 (0.051)</td>
<td>4.138</td>
<td>14.728 (0.002)</td>
<td>23</td>
<td>0.465 (0.107)</td>
<td>-0.041</td>
<td>0.163 (0.921)</td>
<td>7</td>
<td>0.267 (0.113)</td>
<td>0.362</td>
<td>0.980 (0.613)</td>
<td>13</td>
<td>0.912 (0.634)</td>
</tr>
</tbody>
</table>

*Should be between −2.5 and 2.5, item misfit is shown in bold.

bSummarize the deviation of observed from expected responses across sample class intervals. Bonferroni corrected statistically significant (at a level of <0.05) deviations across class intervals, indicating item misfit and bolded in the table.

cShould be non-significant, Bonferroni corrected (each item’s fit to the model).
to less (Figure 2, lower panel). Item locations are covered by the persons, meaning that the basic premises for evaluating the scale are good. However, the range of person locations is not optimally represented by item locations, as illustrated by a relatively large gap between approximately $-0.2$ and $-1.8$ logits, a somewhat smaller gap between about 1 and 2 logits, and non-coverage of persons at the extreme levels of occupational balance (Figure 2, lower panel). This means that the OBQ11 measures with less precision at these intervals.

As illustrated in Table 3, there was support for the logic of the hierarchical item ordering relative to previous research [31,32]. That is, being able to do what one wants is relatively easy to achieve (i.e. does not require very high levels of occupational balance), while experiencing balance between different types of occupations may be more difficult to achieve. For example, item 4 (Time for doing things wanted) represents the lower end of the OBQ11 continuum ($-1.022$ logits) and item 5 (Balance between work, home, family, leisure, rest, and sleep) represents the highest level of occupational balance ($0.378$ logits). When taking measurement uncertainties ($\pm 2$ SEs) into account, one item (number 11) exhibited different locations in the new OBQ11 compared to the original OBQ.

**Discussion**

The present study investigated the validity of the OBQ using RMT resulting in a new version of the...
instrument, the OBQ11. As opposed to previous CTT based analyses [7], several important features were revealed. First, the original response scale did not work as intended and reduction from six to four response categories improved the questionnaire without loss of reliability. Second, two items signaled mis-fit due to multidimensionality, which could be understood theoretically, thus resulting in a conceptually clearer and reliable 11-item questionnaire, with four functioning response categories, acceptable model fit, measurement invariance across age and genders, and probably a logical hierarchical item ordering. The evaluation of the logical hierarchical item ordering is based on two previous studies [31,32] but more research is needed to be sure.

In the analysis of the OBQ, data suggested that item 3 (Perceiving one’s occupations as meaningful) and item 6 (Balance between doing things alone and with others) may represent a construct other than occupational balance. As with all other items, these items were derived from empirical studies. In a qualitative study [8], women with stress-related disorders emphasized the importance of meaningful occupations to experience occupational balance. Furthermore, in a previous literature review [33] one of four identified perspectives of occupational balance was congruence between values and occupations. According to another definition [34, p. 231] occupational balance is ‘the extent to which a pattern of occupation is perceived as harmonious, fulfilling, and compatible with one’s values’. When developing item 3, we, therefore, considered occupations perceived as meaningful to be an indicator of occupational balance. Very likely meaningfulness is affected by the environment and meaning affects occupational balance but is not an indicator of occupational balance. Regarding item 6, women with stress-related disorders have emphasized the importance of balance between doing occupations alone and together with others [8]. However, we have not found any other support for this. Other people are part of the environment and thus not an indicator of occupational balance. As the OBQ11 is a short instrument, focusing on measuring occupational balance from the perspective of people’s experience of their amount and variation of occupations regardless of which these are, it is suitable to combine OBQ11 with measurements of other aspects of occupational balance such as the LBI [2]. Also, combining the OBQ11 with measures of occupational meaning or value, such as the Occupational Value Assessment with predefined items (OVal-pd) [35], and measures of the social environment [36] is likely a feasible approach to capture a wider array of important aspects that affect occupational balance.

The results further show a slight but general tendency for the items to represent less occupational balance than that experienced by the persons, and we found gaps along the occupational balance continuum that are not represented by the OBQ11. This may have practical implications, such as the OBQ11 being less suitable for detecting differences between people with higher occupational balance or to follow changes over time in such samples. However, this is not considered a major problem because precision is of less concern at the highest levels of occupational balance, as those with high occupational balance do not need to improve their occupational balance. Of somewhat more concern are the two gaps in the item coverage, which comprise measurement precision at levels of occupational balance represented by a majority of people. One obvious way to improve precision and targeting would be to increase the number of items and/or response categories to enhance the representation of the latent occupational balance continuum. Given the evidence that a four-category response scale is preferable to six response categories, it seems reasonable to suggest that future efforts should attempt to identify additional items rather than expanding the number of response categories.

A limitation of the present study was that mean age was around 50 and that two thirds were women. This may have affected the results as previous studies have shown that occupational balance differs between men and women [31] between people with and without children [37] and probably between people with disabilities and without them. Therefore, the OBQ11 needs to be tested in additional samples of other ages and a more even gender distribution. Furthermore, it also has to be tested in various patient groups as well as across countries and cultures.

It may also be argued that the sizes of our samples were somewhat limited, particularly regarding sample 1. However, the primary purpose of sample 1 was to get an initial indicator of any major issues related to the OBQ. As such, a sample size of about 150 can be considered as acceptable [38]. As for sample 2, it has been found that samples of about this size represent a fair balance with regard to assessment of fit to the Rasch model using the RUMM2030 software [28].

**Practical implications**

Since occupational balance is an essential aspect of occupational therapy it is vital to be able to measure
it. This study suggests that the OBQ11 may serve to fill this gap by providing a valid and reliable measure of occupational balance. The OBQ11 may have practical relevance as a tool to identify occupational balance in general populations. The OBQ11 can be used as a screening tool to prioritize clients for occupational therapists and for intervention planning in health promoting occupational therapy. In the present study a general population was selected, but the OBQ11 needs to be tested in clinical samples too.

**Conclusions**

The new version of the Occupational Balance Questionnaire with 11 item and four response categories, the OBQ11, satisfies the measurement criteria defined by RMT and may provide a useful means to assess experienced occupational balance both in groups and individuals. However, further studies are needed to validate the OBQ11 across a wider range of applications.

**Acknowledgements**

The authors want to thank all participating respondents, and the occupational therapy students for their valuable help with data gathering.

**Disclosure statement**

The authors report no conflict of interest.

**ORCID**

Carita Håkansson [http://orcid.org/0000-0002-3660-3079]

Peter Hagell [http://orcid.org/0000-0003-2174-372X]

**References**


