Evaluation of the CES-D in Six Countries Using Rasch Item Response Theory (IRT) Analysis

Bushnell DM¹ and Whalley D²

¹Health Research Associates, Seattle, Washington, USA; ²Galen Research, Manchester, UK

OBJECTIVES

To evaluate the scaling properties of the Centers for Epidemiologic Studies – Depression Scale (CES-D) in six countries through the application of the one-parameter (Rasch) IRT model.

METHODS

Data Collection:
Data taken from the Longitudinal Investigation of Depression Outcomes (LIDO). Patients coming for care at participating primary care facilities were invited to complete the screening assessment. Using a cut-point of 16 or above on the CES-D, those who met eligibility criteria (i.e., age 18-75, able and willing to participate in all scheduled visits, provide adequate locator information, willing to sign consent), were enrolled and invited to a baseline visit. The following analysis utilizes this baseline data.

Analysis:
Rasch Item Response Theory provides:
- a method for obtaining objective, fundamental, linear measures from stochastic observations of ordered category responses.
- estimates of item locations (calibrations) along a common measurement continuum expressed in log-odd units (logits) with standard errors.
- information-weighted fit (INFIT) and outlier-sensitive fit (OUTFIT) statistics expressed in mean square (MNSQ) and standardized (ZSTD) values.

RESULTS

Sample:

<table>
<thead>
<tr>
<th>Country</th>
<th>Age (mean (SD))</th>
<th>Gender (% female)</th>
<th>Marital (% married)</th>
<th>Education (mean yrs)</th>
<th>Income* (USD) (median monthly)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Be'er Sheva</td>
<td>41.4 (14.3)</td>
<td>62</td>
<td>67</td>
<td>11.9 (2.7)</td>
<td>542</td>
</tr>
<tr>
<td>Barcelona</td>
<td>41.5 (15.2)</td>
<td>71</td>
<td>56</td>
<td>10.7 (3.4)</td>
<td>564</td>
</tr>
<tr>
<td>Melbourne</td>
<td>38.4 (14.3)</td>
<td>65</td>
<td>29</td>
<td>12.5 (3.8)</td>
<td>481</td>
</tr>
<tr>
<td>Porto Alegre</td>
<td>38.9 (13.6)</td>
<td>75</td>
<td>53</td>
<td>6.3 (3.4)</td>
<td>178</td>
</tr>
<tr>
<td>Seattle</td>
<td>41.8 (14.6)</td>
<td>67</td>
<td>48</td>
<td>13.3 (2.3)</td>
<td>1700</td>
</tr>
<tr>
<td>St. Petersburg</td>
<td>47.0 (16.2)</td>
<td>72</td>
<td>45</td>
<td>11.8 (3.5)</td>
<td>35</td>
</tr>
</tbody>
</table>

Table 2: Items Statistics by country (Logits, Standard Errors, INFIT and OUTFIT values)

Similar patterns were noted across the six countries as shown by the goodness-of-fit statistics (see Table 2). Three items with MNSQ values below the recommended range (<0.85 indicating redundancy) were identical at each site (i.e., I felt depressed, I felt sad, and I could not shake off the blues). Furthermore, items that were above the range (>1.25 indicating that they yield responses that do not conform to the response pattern of most other items) included two of the positively framed items (I felt that I was just as good as other people and I felt hopeful about the future). All items covered the construct well for most language versions, although some gaps in measurement were observed (see Figure 1). Finally, some degree of item overlap/redundancy was evident in all language versions, with several items sharing the same location on the underlying construct.

Table 3: Measurement Continuum by Country

A similar item hierarchy (using item calibrations from Table 2) can be seen across the six countries. The hierarchy goes from the more difficult to endorse items at the top (i.e., items 19 and 15) to the easier to endorse at the bottom (i.e., items 11 and 12).

CONCLUSION

Although the CES-D is a commonly employed measure of depression severity, it may be necessary to take a closer look at its scalability. Such information is vital for the appropriate use and interpretation of multi-center and multi-national data resulting from this instrument.