An item response theory analysis of Wong and Law emotional intelligence scale

Jahanvash Karima *

a CERGAM, Institute d’Administration des Entreprises d’Aix-en-Provence, Glos Cuiot Puyricard-BP 30063, 13089, Aix-en-Provence, France

Received November 2, 2009; revised December 10, 2009; accepted January 18, 2010

Abstract

The purpose of this study was to perform an IRT analysis of Wong and Law Emotional Intelligence Scale (WLEIS: Wong & Law, 2002). The sample consisted of 481 university students in the province of Balochistan, Pakistan. After examining the unidimensionality of WLEIS four sub-scales, graded response model for seven ordered categories was applied (Samejima, 1969) using “ltm” an R package (Rizopoulous, 2006). Results indicated that the WLEIS sub-scales yield precise measurement for individuals with low to moderate trait levels and relatively imprecise measurement for individuals with high trait levels.

Keywords: Item response theory; emotional intelligence; graded response model.

1. Introduction

Emotional intelligence (EI) is a concept that has received a great deal of scholarly attention in the social science literature, as well as, in the popular press. Further to this, meta-analysis results indicate that EI is an important predictor of health-related variables (Schutte, Malouff, Thorsteinsson, Bhullar, & Rooke, 2007) and performance (Van Rooy & Viswesvaran, 2004). Salovey and Mayer (1990) were first to utilize the term “emotional intelligence” to represent the ability to deal with emotions. They drew on relevant evidence from previous intelligence and emotion research and provided the first comprehensive model of EI. Since Salovey and Mayer’s (1990) conceptualization, a considerable amount of theoretical and empirical research has been done on the conceptualization of EI (e.g., Bar-On, 1997; Goleman, 1995; Mayer & Salovey, 1997; Petrides, & Furnham, 2001), as well as, its measures (e.g., Emotional Quotient Inventory: Bar-On, 1997; Mayer-Salovey-Caruso Emotional Intelligence Test: Mayer, Salovey, Caruso, & Sitarenios, 2003; Self-report Emotional Intelligence Test: Schutte et al., 1998; Trait Emotional Intelligence Questionnaire: Petrides, Perez-Gonzalez, & Furnham, 2007).

Despite these notable advances in the EI field, it is an open question whether existing EI instruments possess the requisite psychometric properties. Specifically, the use of EI instruments in selection and promotion has led to a concomitant increase in the need for researchers to evaluate the quality and, more importantly, the measurement precision of the EI instruments. Unlike classical test theory (CTT), item response theory (IRT) based psychometric methods offer some of the best alternatives for devising and optimizing instruments both at test and item level.

* Jahanvash Karim.
E-mail address: j_vash@hotmail.com

1877-0428 © 2010 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.sbspro.2010.03.637
(Embretson & Reise, 2000; Hambleton, 1990; Hambleton, & Rogers, 1990; Szabó, 2008). For instance, in IRT, the model is expressed at the level of the observed item response rather than at the level of the observed test score. Furthermore, unlike CTT which assumes that measurement precision is constant across the entire trait range, IRT models recognize that measurement precision may not be constant for all examinees. Therefore, while evaluating EI instruments, IRT can be useful in detecting/finding the amount of information and precision these EI instruments provide at specific ranges of test score that are of particular interest. For example, for selection purposes measurement precision at the upper end of trait (θ) would likely be of main interest and a lack of measurement precision at the low end of trait (θ) continuum might be excused. In sum, it is likely that many scales used in EI research have an unequal distribution of precision across the normal range of the trait continuum. The purpose of this study is to perform an IRT analysis of one of the widely used self-report measure of trait EI, that is, Wong and Law Emotional Intelligence Scale (WLEIS: Wong & Law, 2002) in a sample of university students.

2. Item response theory (IRT)

In IRT, the underlying trait is commonly designated by Greek letter theta (θ) and is most often scaled to have a mean of zero and standard deviation of one. The correspondence between the responses to an item and latent trait (θ) is known as the item characteristic curve (ICC). ICC is defined as, “the (nonlinear) regression that represents the probability of endorsing an item (or an item response category) as a function of the underlying trait “(Fraley, Waller, & Brennan, 2000. P. 351). Panel A of Figure 1 presents three different ICCs for 3 dichotomous items with options yes and no. As can be seen, the probability of endorsing an item (option) is monotonically nondecreasing in θ, that is, the probability of endorsing an item increases as one moves along the trait continuum (θ). Furthermore, the curves are nonlinear and may differ in shapes.

2.1. One-parameter logistic model (1-PLM)

The simplest commonly used IRT model has one parameter for describing the characteristics of the person and one parameter for describing the characteristics of the item. Consequently each item is supposed to have the same discrimination, which is represented by parallel ICCs as shown in Panel A of Figure 1. This model can be represented by

\[ P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \]

Where \( P_i(\theta) \) is the probability of a random examinee with ability \( \theta \) answering item \( i \) correctly, \( b_i \) is the difficulty parameter for item \( i \), and \( e \) is a natural constant whose value is 2.71182. The item difficulty parameter (b) represents the level of the latent trait necessary to have a 0.50 probability of endorsing the item in the keyed direction. Panel A of Figure 1 is an example of 1-PL model, with three dichotomous items. In 1-PL model only item difficulty is allowed to vary, therefore the three ICCs are parallel (same slopes). The only feature of ICC that changes from test item to test item is the location of the curve on the \( \theta \) scale. For example, if an item has a difficulty value of 1, then an individual with a trait level of 1 has a 50% probability of endorsing item (item 3 in figure 1 panel A).

![Figure 1. Item Characteristic Curves.](image-url)
2.2. Two-parameter logistic model (2-PL)

The two parameter model extends the 1-PL model by estimating an item discrimination parameter ($a_i$) besides item difficulty ($b_i$) (Crocker & Algina, 1986). This model is represented by

$$P_i(\theta) = e^{D(a_i(\theta - b_i))} / [1 + e^{D(a_i(\theta - b_i))}]$$

Where $\theta$ is the ability level, $b_i$ is difficulty parameter, $a_i$ is the discrimination parameter, and D is a scaling factor introduced in order to make the logistic function approximate the normal ogive function as closely as possible. When $D = 1.7$, the value of $P_i(\theta)$ for the 2-PL normal ogive and the 2-PL logistic model are approximately equal (Szabó, 2007).

The item discrimination parameter ($a_i$) represents an item’s ability to differentiate between people with contiguous trait levels. The $a_i$ in IRT is similar to an item-total-correlation of CTT (Fraley et al., 2000; Hays, Morales, & Reise, 2000). Similarly, in IRT the item that has high discrimination value is considered as a better indicator of latent trait. Item 1 in Panel B of Figure 1 has a discrimination value of 1.5. Examinees with trait levels in the vicinity of 1.25 are more likely ($P(\theta) = .60$ approximately) to endorse item 1 than people in the vicinity of 0.75 ($P(\theta) = .36$ approximately). Whereas, item 3 is less steep with discrimination value of 0.50. Hence item three is not efficient in differentiating between people with contiguous trait levels. For instance examinee with the trait level in the vicinity of 1.25 are only slightly more likely to endorse item three ($P(\theta) = .52$ approximately) than are examinees with trait level of 0.75 ($P(\theta) = .48$ approximately). Therefore, we can say that item 3 does a poor job of discriminating among individuals with similar/contiguous trait levels.

We gain information about examinees based on their responses to particular items and the properties of items. The concept of information helps us understand where a given scale provides more or less information about examinees. Information function shows how much psychometric information (a number that represents an item’s ability to differentiate among people) the items provide at each trait level (Fraley et al., 2000; Reise, Ainsworth, & Haviland, 2005). Like item response functions, item information functions can provide useful information about item parameters. Item information in 2-PL is represented by

$$I_j(\theta) = a_j^2 \times P_j(\theta) \times (1 - P_j(\theta))$$

Where $I_j(\theta)$ is item information, $a_j^2$ is the squared item discrimination parameter for item $j$, and $P_j(\theta)$ is the probability of endorsing item $j$ for individuals with $\theta$ level $i$. Panel A Figure 2 shows the item information function corresponding to three items in Panel B of Figure 1. Clearly, items of different difficulty provide information in different-trait ranges, that is, an item is most informative at trait level corresponding to item difficulty values. Moving the threshold ($b_i$) up or down would simply move the IIF right or left on the x-axis. Item 1 and 2 are IIF for two items with different thresholds, but same slope. Furthermore, more discriminating items (e.g., items 1 and 2) provide more information than less discriminating items (e.g., 3). Slopes control how peaked the IIF is. The higher the slope values, the more information that item provides around threshold. To understand how the test is functioning as a whole, item information can be summed to produce a scale information function. The height of TIF is proportional to the standard error of measurement (SEM). The square root of the inverse of information at a given level of the latent trait provides the standard error which would be attached to that particular score. Items and scale information functions are analogous to CTT’s item and test reliability. However, under IRT framework information (measurement precision) can potentially differ for people with different trait level, whereas in CTT the scale reliability (precision) is the same for all individuals regardless of their raw score levels. In sum, in IRT there are as many standard errors of measurement as there are unique trait estimates (Fraley et al., 2000).
2.3. The graded response model (GRM)

The graded response model (GRM: Samejima, 1969, 1997), an extension of the 2-PL model, is appropriate to use when item responses can be characterized as ordered categorical responses such as exists in Likert rating scales. In the GRM, each scale item \( i \) is described by one item slope parameter \( a_i \) and \( j = 1 \ldots m_i \) between category threshold parameters \( b_{ij} \). Each threshold specifies the point on the \( \theta \) scale at which a subject has .50 probability of responding in some higher category than the one to which the threshold belongs. These thresholds share a common slope or discrimination factor, \( a_i \). For instance, for an item where examinee receive item score of 0, 1, 2, 3, & 4 (5 options), there are \( m_i = 4 \) thresholds \((j = 1 \ldots 4)\) between the response options. GRM let us determine the location of these thresholds on the latent trait continuum. There are mainly two steps involved in computing the response probabilities in the GRM: computation of \( m_i \) curves (or option/operating characteristic curves) for each item and then computing the actual probability of endorsing a particular response option. For the GRM, one operating characteristic curve needs to be estimated for each between category thresholds. As already noted above, an item response scale is conceptualized as a series of \( m-1 \) response dichotomies, where \( m \) represents the number of response options for a given item. An item related on a 1-to-5 has four response dichotomies: (a) category 1 versus categories 2, 3, 4, and 5; (b) categories 1 and 2 versus categories 3, 4, and 5; (c) categories 1, 2, and 3 versus categories 4, and 5; and (d) categories 1, 2, 3, and 4 versus category 5. In GRM, 2-PL model are estimated for each dichotomy with the constraints that the slopes of each of the operating characteristic curves are equal within an item.

In the second step, the operating characteristic curves for each response dichotomy are used to calculate the probability of endorsing a particular response option, \( x_j \), as a function of the latent trait. Embretson and Reise (2000) called these probability functions as category response curves (CRC). The CRC for a particular response option, \( x_j \) is given by the following equation

\[
P_i(x_j) = P_{x_j}(\theta) - P_{x_j+1}(\theta)
\]

Where \( P_{x_j}(\theta) \) is the probability of endorsing option \( x_j \) or higher and \( P_{x_j+1}(\theta) \) is the probability of endorsing a next highest option, \( x_j+1 \), or higher. By definition, the probability of responding in or above the lowest category is \( P_{x_0}(\theta) = 1 \), and the probability of responding above the highest category is \( P_{x_m}(\theta) = 0 \). Thus with a 5-point scale, the probability of responding in each of the five categories (CRCs) are given as follows

\[
P_{i0}(\theta) = 1.0 - P_{x_1}(\theta)
\]
\[
P_{i1}(\theta) = P_{x_1}(\theta) - P_{x_2}(\theta)
\]
\[
P_{i2}(\theta) = P_{x_2}(\theta) - P_{x_3}(\theta)
\]
\[
P_{i3}(\theta) = P_{x_3}(\theta) - P_{x_4}(\theta)
\]
\[
P_{i4}(\theta) = P_{x_4}(\theta) - 0
\]

To better illustrate these points, operating characteristic curves and CRCs for a 5-point item that fits the GRM are given in Panel A & B of Figure 3. The item has a discrimination value \( (a_i) \) of 1.5 and difficulty or threshold values \( (b_{im}) \) of -1.5, -0.5, 0.5, and 1.5. From Figure it is evident that the between category threshold parameters represent the point along the latent trait scale at which examinees have a .50 probability of endorsing in or above category. Panel B, Figure 2 shows the CRCs for this item. These curves represent the probability of responding in each category \((x = 0, \ldots, 4)\) conditional on examinee trait level and for any fixed level trait, the sum of the response probabilities is equal to 1. The shape and location of the operating characteristic curves and category response curves depends upon the item parameters. For instance the higher the \( a_i \) (slope parameter) the steeper the operating characteristic curve and more peaked and narrow the CRCs, indicating that the response categories differentiate among trait levels fairly well. Generally speaking, items with higher shape and parameters provide more information.

The threshold parameters \( (b_{ij}) \) determine the location of the curves and where each of the CRC (Figure 3 Panel B) from the middle response options peaks (i.e., middle of two adjacent threshold parameters). In ploytomous IRT models, these values should not be interpreted directly as item discrimination. To directly assume the amount of discrimination item values provide a researcher need to compute item information curves (IICs). Item information functions can be generated for graded response items. Equations for the GRM information functions are provided in Samejima (1969, p. 39).
3. Method

3.1. Participants

The sample consisted of 481 university students (233 males, 246 females, 2 unreported), ranging in age from 20 to 45 years ($M = 28.12$ years, $SD = 8.7$). The participants were obtained from three universities in the province of Balochistan, Pakistan via non-probability purposive sampling (Cohen, Manion, & Morrison, 2000, p.99). All participants were treated in accordance with the “Ethical principles of Psychologists and Code of Conduct” (American Psychological Association, 2002). Administration of the questionnaires was carried out by post graduate students who acted as research assistants and no monetary incentive was provided.

3.2. Instruments

Wong and Law Emotional Intelligence Scale (WLEIS: Wong & Law, 2002). WLEIS consists of 16 items and taps individuals’ knowledge about their own emotional abilities rather than their actual capacities. Specifically, the WLEIS is a measure of beliefs concerning self-emotional appraisal (SEA) (e.g., “I have a good sense of why I have certain feelings most of the time”), others’ emotional appraisal (OEA) (e.g., “I always know my friends’ emotions from their behavior”), regulation of emotion (ROE) (e.g., “I always set goals for myself and then try my best to achieve them”), and use of emotion (UOE) (e.g., “I am able to control my temper and handle difficulties rationally”). The response scale has been seven point Likert-type scale ranging from one (strongly disagree) to seven (strongly agree). Preliminary psychometric analysis (i.e., reliability, factorial, discriminant, convergent, and predictive validity) of the WLEIS suggests that this scale is reliable and valid self-report index of the ability to monitor and manage emotions (e.g., Law, Wong & Song, 2004; Shi & Wang, 2007; Wong & Law, 2002). Coefficients alphas for the four dimensions were: SEA: .82; OEA: .80; ROE: .79; UOE: .78.

3.3. Analysis

IRT models assume that the latent trait construct space is either strictly unidimensional, or as a practical matter, demonstrated by a general underlying factor. Since WLEIS is composed of four different underlying factors (i.e., SEA, OEA, UOE, ROE), each scale had to be examined separately for both unidimensionality and item response characteristics. In IRT framework assessment of unidimensionality is often done using the DIMTEST (Stout, 1990). However, the WLEIS subscales do not have enough items per scale (only 4) to enable valid application of DIMTEST. As an alternative, unidimensionality was assessed by examining the relative ratio of the eigenvalues of the first and second factor. Principal axis factor analysis on the items polychoric matrix was conducted. The presence of unidimensionality in the data is supported if the loadings on the second factor is < .30 (Roberson-Nay, Strong, Nay, Beidel, & Turner, 2007) and the variance accounted for (VAF) by the first factor should be above the 0.40 threshold (Smith & Reise, 1998).

A graded response model for seven ordered categories was applied (Samejima, 1969) using “ltm” an R package (Rizopoulous, 2006). ltm generates item parameters and their standard errors along with item and total scale information at various levels of the $\theta$. The maximal number of iterations (cycles) was set to 2000, and all scales converged before reaching 2000. There are many methods for assessing model-data fit in IRT analysis. In this study
fit of parameters obtained for each scale was evaluated using the graphical procedure. Fit plots are most widely used method for examining model-data fit. Fit plots for all options associated with 16 WLEIS items were computed using MODFIT (Stark, 2002). MODFIT computes two functions, that is, theoretical item response function (IRT), and empirical item response function (EMP), with 95% confidence intervals for the empirical points. A close correspondence between the IRF and EMP curves suggests that the model fits the data well.

4. Results

4.1. Unidimensionality

Unidimensionality of each WLEIS scale was assessed by principal axis factoring applied to a matrix of polychoric correlations (used for items with ordered responses) with Stata (2005). Table 1 presents statistics relevant to scales dimensionality: Internal consistency using CTT method was reasonably strong ranging from .78 for the UOE scale to .82 for the SEA scale. As can be seen, there was prominent first factor as indicated by eigenvalues. All items had factor loadings of greater than 0.70 for every scale on the first factor (results not presented). Furthermore, the variance accounted for by the first factor was well above the .40 threshold recommended by Smith and Reise (1998). These results provide support for unidimensionality of each WLEIS sub-scale.

Table 1. Descriptive statistics relevant to dimensionality of WLEIS sub-scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Cronbach’s Coefficient Alpha (raw)</th>
<th>First Eigenvalue</th>
<th>Variance Explained by First Eigenvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEA</td>
<td>.82</td>
<td>2.70</td>
<td>67.49%</td>
</tr>
<tr>
<td>OEA</td>
<td>.80</td>
<td>2.65</td>
<td>66.42%</td>
</tr>
<tr>
<td>UOE</td>
<td>.78</td>
<td>2.62</td>
<td>66.52%</td>
</tr>
<tr>
<td>ROE</td>
<td>.79</td>
<td>2.67</td>
<td>65.92%</td>
</tr>
</tbody>
</table>

4.2. Parameter estimation

Four separate graded response models were calibrated for each of the WLEIS scale. The graded response models were successfully calibrated by “ltm” (R package) for each WLEIS scale with a maximum intercycle parameter change of less than 0.012 indicating that marginal maximum likelihood algorithm had converged (Thissen 1991).

In order to check the appropriability of the unconstrained GRM models, constrained version of the GRM was compared with the unconstrained model through likelihood ratio tests. In contrast to unconstrained models, constrained version of GRM assumes equal discrimination parameters across all set of items present in the test (here four items within each WLEIS sub-scale). Likelihood ratio test revealed that the unconstrained GRMs provide better fit than the constrained GRMs for all WLEIS scales (LRT\text{SEA} = 40.24, p < .001; LRT\text{OEA} = 24.71, p < .001; LRT\text{UOE} = 26.3, p < .001; LRT\text{ROE} = 27.01, p < .001).

A summary of the discrimination and threshold item parameters for unconstrained GRMs is given in the Table 2. Results indicated considerable variation in the $a$, discrimination parameter across the items within each WLEIS sub-scale (SEA: 1.41 – 3.18; OEA: 1.55 – 2.36; UOE: 1.78 – 2.18; and ROE: 1.42 – 2.53). Table 2 also revealed that the category threshold values were somewhat skewed toward the negative range of $\theta$.

Table 2. Item parameters

<table>
<thead>
<tr>
<th></th>
<th>$a$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
<th>$b_5$</th>
<th>$b_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEA1</td>
<td>2.03</td>
<td>-2.59</td>
<td>-2.05</td>
<td>-1.51</td>
<td>-1.11</td>
<td>-0.42</td>
<td>0.68</td>
</tr>
<tr>
<td>SEA2</td>
<td>3.18</td>
<td>-2.96</td>
<td>-1.94</td>
<td>-1.51</td>
<td>-1.23</td>
<td>-0.68</td>
<td>0.40</td>
</tr>
<tr>
<td>SEA3</td>
<td>2.64</td>
<td>-2.63</td>
<td>-2.15</td>
<td>-1.90</td>
<td>-1.52</td>
<td>-0.63</td>
<td>0.31</td>
</tr>
<tr>
<td>SEA4</td>
<td>1.41</td>
<td>-3.98</td>
<td>-3.08</td>
<td>-2.35</td>
<td>-1.43</td>
<td>-0.54</td>
<td>0.66</td>
</tr>
<tr>
<td>OEA1</td>
<td>1.98</td>
<td>-2.83</td>
<td>-2.33</td>
<td>-1.62</td>
<td>-1.20</td>
<td>-0.41</td>
<td>0.76</td>
</tr>
<tr>
<td>OEA2</td>
<td>2.36</td>
<td>-2.82</td>
<td>-1.83</td>
<td>-1.57</td>
<td>-1.01</td>
<td>-0.17</td>
<td>0.73</td>
</tr>
</tbody>
</table>
Category response curves were computed for each item. Response probabilities between 0 and 1 were plotted over the ±4 standard deviation range on the \( \theta \) continuum. Response categories are ordered from 1 to 7 from the negative to positive range of \( \theta \). Inspection of option characteristic curves revealed more or less consistent ordering of category responses as a function of \( \theta \) (almost for all items). As expected, the probability of endorsing option 7 (“Strongly Agree”) increases as \( \theta \) increases and the probability of endorsing option 1 (“Strongly Disagree”) decreases as \( \theta \) increases (Appendix 1).

Figure 4 displays the item information and test information functions for the WLEIS four subscales. These IIFs and TIFs indicate the area on the \( \theta \) continuum in which the WLEIS items and scales provide the most information or best discrimination among the examinees.

As can be seen in these IIF plots, few of the curves (i.e., SEA4, OEA3, and ROE3) are relatively low. This indicates that overall degrees of measurement precision for these items are also relatively low. As all WLEIS scales had uneven distribution of item difficulty threshold values (Table 2), therefore it was expected that they will have an uneven information function. Figure 4 also displays the test information function for the four WLEIS scales. As can be seen in these plots, all scales lack uniform measurement precision across wide regions of their respective trait ranges. Typically, the scales are less precise for measuring individuals with theta levels falling above 1.00 and below -2. In sum, the WLEIS scales yield precise measurement for individuals with low to moderate trait levels and relatively imprecise measurement for individuals with high trait levels (Table 3).

Fit plots for 4 GRM models were computed using Modfit. Because of the prohibitively large number of plots (4 scales * 4 items * 7 response categories = 132 plots) only fit plots for responses associated with SEA1 are presented here (Appendix 1). Responses associated with WLEIS 16 items depicted more or less the same kind of fit plots. Inspection of these fit plots revealed close correspondence between IRF (Item response function computed from calibration sample) and EMP (empirical item response function computed from a cross-validation sample). The results suggest that the GRM model for the WLEIS scales fits the data well.

### Table 3. Information within different theta ranges

<table>
<thead>
<tr>
<th></th>
<th>(-4, 4)</th>
<th>(-4, 0)</th>
<th>(0, 4)</th>
<th>(-4, 1)</th>
<th>(-2, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEA</td>
<td>30.26</td>
<td>21.85 (72.2%)</td>
<td>7.48 (24.71%)</td>
<td>27.29 (90.2%)</td>
<td>19.12 (63.19%)</td>
</tr>
<tr>
<td>OEA</td>
<td>29.41</td>
<td>20.17 (68.56%)</td>
<td>8.57 (29.15%)</td>
<td>25.54 (86.84%)</td>
<td>17.60 (59.84%)</td>
</tr>
<tr>
<td>UOE</td>
<td>25.71</td>
<td>18.05 (70.22%)</td>
<td>6.63 (25.78%)</td>
<td>22.31 (86.77%)</td>
<td>14.36 (55.87%)</td>
</tr>
<tr>
<td>ROE</td>
<td>28.01</td>
<td>18.01 (64.25%)</td>
<td>9.77 (34.87%)</td>
<td>23.85 (85.13%)</td>
<td>19.12 (63.19%)</td>
</tr>
</tbody>
</table>
5. Discussion

The purpose of this research was to assess and evaluate the precision and accuracy of Wong and Law Emotional Intelligence scale (WLEIS: Wong & Law, 2002) across the whole range of emotional intelligence continuum, in order to determine how useful the WLEIS is for selection, promotion, training and development purposes. The study demonstrated that IRT modeling for graded responses (Samejima, 1969, 1997) is a useful approach to item analysis with the WLEIS. The IRT (GRM) treats the response categories at the appropriate ordinal level of data, unlike typical CTT which imposes continuous level assumptions on ordered responses (Uttaro & Lehman, 1999). The results of this study provided important details about WLEIS’s items and scales performance through item parameter estimations, IIFs, and TIFs. The results of this study supported the idea that unconstrained GRM is preferable for the WLEIS than constrained GRM (that assumes equal discrimination parameter across items). Examination of the item parameters indicated considerable variation in the discrimination parameter across all the items within each WLEIS subscale. Furthermore, category threshold values \( b_i \) were mostly skewed toward the negative range of \( \theta \). In other words, the item threshold values were concentrated in a narrow region of the trait range, that is, toward negative end.

Inspection of item information curves revealed that the overall degree of measurement precision for SEA4, OEA3, and ROE3 were relatively low. Furthermore, inspection of both IIF and TIF revealed that item discrimination values were concentrated in certain regions of the trait range and had uneven information functions. IIFs and TIFs for 16 items and 4 scales revealed that WLEIS performs well for respondents with low to moderate levels of emotional intelligent ability. Specifically, TIFs for all WLEIS scales provided maximum information in the range of -2 to 1 on trait continuum and become increasing unreliable in assessing the respondents with high levels of emotional intelligence ability.

As with most personality instruments, measurement precision tends to decline somewhat at the extreme ends of the latent traits. Results of this study indicated that WLEIS is not able to discriminate among people with high EI, that is, it does not seem to be appropriate EI instrument for selection or promotion of individuals high on EI where the focus is on the higher level of EI (\( \theta \)). However, WLEIS seems to be suitable for screening out individuals who have low to moderate levels of EI. Furthermore, the results of this study indicated few items with low information (i.e., SEA4, OEA3, and ROE3). WLEIS scales could be improved by removing such kind of low informative items with low discrimination power, and/or by adding new and relatively difficult items located at the higher end of the EI continuum.
References


Appendix 1

Category response curves for WLEIS 16 items

Example of fit plots for SEA1