Validation of the Brief Version of the Recovery Self-Assessment (RSA-B) Using Rasch Measurement Theory

Skye P. Barbic, Sean A. Kidd, Larry Davidson, Kwame McKenzie, and Maria J. O’Connell


CITATION
Validation of the Brief Version of the Recovery Self-Assessment (RSA-B) Using Rasch Measurement Theory

Skye P. Barbic
University of British Columbia and Social Aetiology of Mental Illness Program, Canadian Institute of Health Research Initiative, Centre for Addiction and Mental Health, Toronto, Canada

Sean A. Kidd
University of Toronto, and Centre for Addiction and Mental Health, Toronto, Ontario, Canada

Larry Davidson
Yale University School of Medicine

Kwame McKenzie
University of Toronto and Centre for Addiction and Mental Health, Toronto, Canada

Maria J. O’Connell
Yale University School of Medicine

Objective: In psychiatry, the recovery paradigm is increasingly identified as the overarching framework for service provision. Currently, the Recovery Self-Assessment (RSA), a 36-item rating scale, is commonly used to assess the uptake of a recovery orientation in clinical services. However, the consumer version of the RSA has been found challenging to complete because of length and the reading level required. In response to this feedback, a brief 12-item version of the RSA was developed (RSA-B). This article describes the development of the modified instrument and the application of traditional psychometric analysis and Rasch Measurement Theory to test the psychometrics properties of the RSA-B.

Methods: Data from a multisite study of adults with serious mental illnesses (n = 1256) who were followed by assertive community treatment teams were examined for reliability, clinical meaning, targeting, response categories, model fit, reliability, dependency, and raw interval-level measurement. Analyses were performed using the Rasch Unidimensional Measurement Model (RUMM 2030).

Results: Adequate fit to the Rasch model was observed ($\chi^2 = 112.46, df = 90, p = .06$) and internal consistency was good ($r = .86$). However, Rasch analysis revealed limitations of the 12-item version, with items covering only 39% of the targeted theoretical continuum, 2 misfitting items, and strong evidence for the 5 option response categories not working as intended.

Conclusions: This study revealed areas for improvement in the shortened version of the 12-item RSA-B. A revisit of the conceptual model and original 36-item rating scale is encouraged to select items that will help practitioners and researchers measure the full range of recovery orientation.

Keywords: Recovery, Rasch measurement theory, mental health, validity

Health providers have increasingly come to grapple with the problems that surround transforming services for persons with psychiatric disabilities from custodial modes of service provision to those that facilitate recovery (Davidson, O’Connell, Tondora, Styron, & Kangas, 2006; Davidson et al., 2007). In the context of psychosocial rehabilitation, recovery is defined as “living a satisfying, hopeful, and contributing life even with limitations caused by the illness” (Anthony, 1993). A construct that has emerged from this definition is recovery-orientated care, which in turn is described as a system of support that enables clients to pursue their own meaningful goals by providing care that is individually tailored, strength-based, respectful of rights, and promotes consumer involvement and hope (Kidd, McKenzie, Collins, et al., 2014; Kidd, McKenzie, & Virdee, 2014;...
O’Connell, Tondora, Croog, Evans, & Davidson, 2005). A crucial part of this process of system change is the development of assessment tools that can facilitate and inform change in recovery-oriented services and systems.

Currently, there exist a range of tools, or rating scales, that evaluate the extent to which a system or program meets the principles of recovery. These self-reported rating scales are often filled out from the perspectives of persons in recovery (Shanks et al., 2013; Sklar, Groessl, O’Connell, Davidson, & Aarons, 2013), consultants (Cambell-Orde, Chamberline, Carpenter, & Leff, 2005) and service providers (Cambell-Orde et al., 2005; Salyers, Brennan, & Kean, 2013; Williams et al., 2012). Increasingly, the information gleaned from these rating scales is of central importance for decisions made in routine clinical practice, research, and clinical trials (Cano & Hobart, 2011; J. C. Hobart, Cano, Zajicek, & Thompson, 2007; Kidd et al., 2010; Leucht, 2014; Salyers, Tsai, & Stultz, 2007; Williams et al., 2012). Thus, rating scales used to measure recovery-orientation need to be fit for purpose (i.e., reflect the concept of interest in the context of use), clinically interpretable, and scientifically robust.

The Recovery Self Assessment (RSA; O’Connell et al., 2005) is among the most widely used rating scales to facilitate reflection on the strengths and limitations of services within a recovery framework. This 36-item questionnaire has versions for administrators, service providers, family members/key supports, and clients. Items cover five domains: Life goals versus symptom management; Consumer involvement and recovery education; Diversity of treatment options; Rights and respect; and Individually tailored services (O’Connell et al., 2005). The RSA allows for a generation of a total mean score, domain means, and for the comparison of stakeholder perspectives (Kidd et al., 2011). The consumer version of the RSA assesses the perceptions of individuals with lived experience about whether the system and its providers embrace the core principles of recovery.

The RSA has undergone varying degrees of traditional psychometric testing including examination of internal consistency, test–retest reliability, content, convergent, and discriminant validity (O’Connell et al., 2005; Salyers et al., 2007). While having the above strengths, an area for further development has been identified. It had been suggested that the “person in recovery” version of the tool in its original 36 item format was, for some, challenging to complete because of length of the survey and the level of reading required (Kidd et al., 2011; Kidd et al., 2010). This was identified as a challenge for consumers who were less well or had less formal education. In response to this feedback, a brief 12-item version of the person in recovery RSA was developed (RSA-B) for use in Assertive Community Treatment (ACT) settings.

In the development of the RSA-B, 12 items were identified that allowed for representation of all of the 5 RSA domains. These items were identified through high loadings in factor analyses in the initial validation study of the RSA and expert consensus. Following the identification of these 12 items, the survey development team worked collaboratively to make minor changes to item wording to increase clarity and to specify that ACT services are being evaluated. All items were rated on a 5-point scale (strongly disagree to strongly agree) as was the case with the original RSA, and items were ordered such that domains were evenly dispersed across the questionnaire. The next step was to test the psychometric properties of the shortened version.

Approaches to testing the psychometric properties of rating scales used in mental health commonly use traditional paradigms of testing such as classical test theory (CTT) (Novick, 1966; Traub, 1997). One of the limitations of CTT is that the information generated is sample specific (Cano & Hobart, 2011; Streiner & Norman, 2007). This makes comparison across groups or individuals very difficult. To address this limitation, application of additional psychometric testing methods is encouraged (Cano & Hobart, 2011; Cook et al., 2007; Hobart & Cano, 2009; Hobart et al., 2012; van den Berg, Paap, & Derks, 2013). One example are methods guided by Rasch Measurement Theory (Andrich, 2011).

Rasch measurement methods are used to measure the extent to which the observed rating scale data (in this instance, the clients’ ratings on items in the RSA-B) “fit” with predictions of those ratings from the Rasch model (which defines how a set of items should perform to generate reliable and valid measurements; Andrich, 2011; Cano & Hobart, 2011; Cano et al., 2014). The difference between the expected and observed scores reveals the extent to which valid measurement is achieved (Cano et al., 2014). Compared with traditional methods commonly used, Rasch measurement methods reveal clinically important information about how well items in a rating scale cover the full range of a concept of interest (in this case recovery orientation) and are appropriate for the context of use (in this case people who use Assertive Community Treatment [ACT] services). In addition to ensuring scales fulfill the mandated traditional psychometric criteria, Rasch methods provide important detailed diagnostics for scale improvements and clinical interpretation (Cano, Barrett, Zajicek, & Hobart, 2011; Cano & Hobart, 2011).

The global aim of this study was to provide clinicians and researchers with a comprehensive psychometric evaluation of the RSA-B for people with serious mental illness using traditional and Rasch measurement methods. The specific aims were to (a) evaluate the measurement properties for the RSA-B in people with serious mental illness who use ACT services to enable researchers and clinicians to judge its performance, (b) compare and contrast traditional and Rasch-based psychometric evidence for the reliability and validity of the RSA-B in people with serious mental illness, and (c) help researchers and clinicians increase their awareness of how modern methods can inform decision making in psychiatric rehabilitation.

**Method**

**Participants**

The findings of this study form a part of a larger study examining fidelity of Ontario community mental health teams to the ACT model and functional outcomes of ACT clients (Kidd et al., 2010). All aspects of this study were reviewed by an Institutional Ethics Review Board. A total of 67 ACT teams participated in the study for a total of 79 teams (85% response rate). Reasons for nonparticipation revolved primarily around either staff/coordinator turnover or recentness of start-up. Teams ranged in years of operation from 1 to 17 years (average 6.54 years).

**Procedures**

A package containing detailed instructions, consent forms, participant reimbursement, and rating scales (including the RSA-B)
was mailed to all ACT teams in the province of Ontario. Return envelopes were also provided and each team was instructed to send all materials back to the research coordination center. With respect to the RSA-B, peer support staff were requested to approach 30 clients based on consecutive visits (to avoid purposeful selection of survey-amendable clients). Once permission was obtained from consumers to participate in the study, peer support staff provided clients with the RSA-B and an envelope to put it in when completed.

Client participants received a $5 honorarium for their time. A total of 1256 ACT clients, all adults experiencing severe and persistent mental illness, completed the RSA-B (response rate 63% of target sample). The mean number of clients who participated per team was 21 (range 4–30), representing an average of 35% of the total number of ACT clients per team at the time of study.

**Measure**

The RSA-B consisted of 12 items, each scored on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree), with a higher score indicating greater recovery orientation (see Appendix for the original pool of candidate items).

**Traditional Psychometric Testing of the RSA-B**

The first steps in psychometric testing included correlational and descriptive analyses to evaluate scaling assumptions (legitimacy of summing items), reliability, and validity. RSA-B data were examined for data quality (percent missing data for each item), scaling assumptions (similarity of item means and variances; magnitude and similarity of corrected item-total correlations), scale-to-sample targeting (score means and standard deviations; floor and ceiling effects), and internal consistency reliability (Cronbach’s alpha).

**Rasch Measurement Testing of the RSA-B**

We used Rasch measurement methods to contribute evidence toward how well the shortened 12-item set can capture the full range of the concept of interest (recovery-orientation) in this context of use (Assertive Community Treatment). Rasch measurement methods have been described in great depth elsewhere (Cano et al., 2011; Cano & Hobart, 2011; Pallant & Tennant, 2007; Tennant, 2011; Tennant & Conaghan, 2007). In brief, a scale that defines the full spectrum of a construct will range from −4 to +4 logits, corresponding to ±4 standard deviations defining the full range of a standard normal distribution (low to high recovery orientation). The Rasch model is a probabilistic model that expresses the probability of an item that represents a given level of the construct (in our case level of recovery orientation) being passed (or agreed with) by people with a given level of ability (or perception of recovery orientation from low to high) as a logistic function of the difference between item difficulty and person ability (Rasch, 1980).

In Rasch measurement, two choices of parameterization are available: (a) the partial credit model and (b) the rating scale model. The rating scale model specifies that a set of items share the same rating scale structure (Massof, 2012; Meyer & Hailey, 2012). The partial credit model specifies that each item has its own rating scale structure (Yamada, 2000). The partial credit model derives from multiple-choice tests where responses that are incorrect, but indicate some knowledge, are given partial credit toward a correct response (Massof, 2012; Yamada, 2000). In previous health outcomes research of secondary data, the amount of partial correctness has been found to vary across items (Barbic, Bartlett, & Mayo, in press; Covic, Pallant, Conaghan, & Tennant, 2007; Pallant & Tennant, 2007). Given this was a secondary analysis and we were unaware of what the thresholds looked like in advance of data collection or analysis, we hypothesized that the amount of partial correctness would vary across time. As a result, we used the Masters’ partial credit model (Masters, 1982) to guide a series of iterative steps to measure the degree to which rigorous measurement criteria were achieved. This was done using five key tests: fit, targeting, dependency, multidimensionality, and reliability (Cano et al., 2014). All analyses were performed using Rasch Unidimensional Measurement Model (RUMM2030) software (Andrich, Sheridan, & Luo, 2010).

**Fit.** To measure the extent to which items in the RSA-B work together to capture the recovery-orientation of individuals receiving ACT services, we tested the performance of each item by visually inspecting the (a) ordering of response options, (b) ordering of item thresholds, (c) two statistical indicators (inspection of the individual item fit residuals (±2.5) and chi square), and (d) one graphical indicator (visual inspection of the item characteristics curves).

**Targeting.** We carefully looked at how people and items were distributed along the proposed latent recovery-orientation continuum. We also looked at how well the 12 items covered the full range of the continuum (±4 logits) and targeted the sample under investigation. In other words, we gauged the calibration of the instrument to the population by comparing graphically how closely the amount of recovery orientation displayed by the respondents was adequately measured by the items on the scale (Wright, 1982). We also flagged items in similar locations as potentially redundant and warranting further investigation.

**Dependency.** Dependency refers to the extent to which the response of any of the items in each scale are directly influenced by the response to any other item on the scale (Marais & Andrich, 2008). We examined the residual item correlations and ruled out dependency if correlations were <0.3 (Smith, 2000).

**Multidimensionality.** Given that the original theoretical work highlighted that the construct of instruct covered five domains, we tested the extent to which items from these domains could collectively measure a single latent construct (recovery orientation). We did this by examining how items loaded onto the first principal component of residual scores (Smith, 2002). In the event that two subgroups of items were formed (i.e., one comprising all the highest positive loading items and another for all the lowest negative loading items), we used Smith independent t tests (Smith, 2002) to assess whether person estimates derived from the subtests of items were significantly different from each other. If more than 5% of t tests were significant, explanation for the anomaly was put into question (Andrich, 2011; Smith, 2002).

**Reliability.** We assessed reliability using the Person Separation Index (PSI) (Andrich, 1982), which is analogous to the Cronbach’s alpha (Cronbach, 1951). A value of 0.70 and above was
considered acceptable as an indicator for group use, and 0.70 through 0.85 for individual use (Andrich, 1982).

Results

Traditional Analysis

Initial analysis revealed that data quality was high. Scale scores were computable for 98% of respondents, scaling assumptions were verified (similar mean item scores, and scale scores spanned the measurement continuum), and internal consistency was good (Cronbach’s alpha 0.86). However, as shown in Table 1, there was evidence that scores were notably skewed, with mean scores near the upper end of the scale. Ceiling effects were also noted, with 10% of the sample reporting a maximum score.

Rasch Measurement Analysis

Fit. All 12 items were found to have problems with fit. All items were found to have reversed thresholds, indicating that the 5 response categories were not working as intended. More specifically, there was strong evidence that the ordinal numbering of the categories (i.e., 1 = strongly disagree, 5 = strongly agree) did not accord to their substantive meaning (i.e., a lower score quantifying less recovery-orientation). Figure 1 shows an example of an item with that disordered thresholds. This figure shows that participants were unable to differentiate between the middle scoring options on all items; specifically, participants were unable to distinguish between the middle response categories “2,” “3,” and “4” (labeled disagree, neither agree nor disagree and agree, respectively). As a result, we rescored each item from five response category options to three.

Figure 2 shows the new ordering of the threshold categories for each item. The item maps specifically shows a person’s expected score to each item as a function of the measure of recovery-orientation. The x axis represents the theoretical continuum of the latent construct (less to more recovery orientation). The y axis lists the items included in the final. In this case, the five item response categories were collapsed in to three response category. Figure 2 depicts response categories that are ordered and working as intended, suggesting a three category response option may be more favorable for this sample under study. As a result, for the remaining parts of the analysis, we interpreted the data from the set of item that were rescored.

After rescoring, we found strong statistical and graphical evidence for misfit of two items. Both items had fit residuals outside the −2.5/+2.5 range and both had chi-square probabilities of less than 0.01. These items were: “ACT staff do not threaten, bribe, or force me to do things that I don’t want to do” (item 1 fit residual = 12.62; \( \chi^2 = 669.59, df = 9, p < .01 \)) and “ACT staff members address my personal life experiences, interests, and needs, and also address my unique culture” (item 11: fit residual = −3.58; \( \chi^2 = 100.39, df = 9, p < .01 \)). Both items also had ICCs below the theoretical curve, providing evidence of poor discrimination ability. This suggests that these two items do not fit ideally with the intended construct of the scale. Item fit for the remaining 10 items is reported in Table 2. Table 3 summarizes the model fit with and without the two misfitting items. Table 3 shows that overall fit to the Rasch model improves when these two items are removed, specifically showing a notable decrease in the chi-value suggesting that model fit was not achieved by chance alone.

Targeting. Figure 3 shows the targeting of the sample to the 10 items remaining. Our analysis suggests that the items in RSA-B do not adequately target the sample under investigation, capturing only 54% of the sample. Specifically, items (shown as bars above the x axis in Figure 3) included in the shortened scale do not adequately capture the people (depicted by the bars below the x axis Figure 3) who report higher levels of recovery-orientation. Figure 3 shows dark arrows identifying clear gaps along the measurement continuum where the concept of interest is not being well captured in this context of use.

Dependency. No items had a residual correlation larger than 0.3, showing little evidence of response dependency between any items. This indicates that there is little evidence that the response on one item influences how a person responds to another item.

Multidimensionality. Examination of the Eigenvalues from the principal component analysis suggested the presence of two subscales. This was also supported by the loadings in the first principle component that showed clear patterns of residuals on two components, with 3 items with positive correlations, and 3 others with negative loadings. However, evidence from the t test grouping these items together in subtests revealed that the amount of multidimensionality was not significant, with <1% of the subtests (n = 13) showing no significant differences in the estimated differences generated (t = −3.23, p = .15).

Reliability. Scale reliability was marginally acceptable (PSI = 0.69; standard cut-off for acceptability is 0.70), indicating the items may not adequately separate this sample along the measurement continuum.

Discussion

In the last decade, there has been an increasing demand for accountability of a broad range of mental health services to be recovery-oriented. The RSA-B 12-item scale was intended for
use as an efficient tool in community practice to measure the construct of recovery-orientation for people who utilize mental health services. The global aim of this study was to provide clinicians and researchers with a comprehensive psychometric evaluation of the RSA-B for people with serious mental illness. Although previous studies using traditional test theory techniques have found support for reliability and validity of the 36-item version of the RSA and aspects of the RSA-B, to our knowledge, this is the first study to systematically assess a shortened version of the scale using both traditional and Rasch psychometric methods.

The specific aims of this study were to provide a comprehensive psychometric evaluation of the RSA-B for people with serious mental illness using both traditional and Rasch measurement methods and to compare the use of both methods to inform research and clinical practice. We found that the shortened version of the RSA-B, in its current format, does not meet the criteria for robust measurement of recovery orientation for this population. Although internal consistency was good and data quality was high, traditional analysis suggested evidence that the items were behaving aberrantly. Rasch analysis uncovered a detailed profile of the

![Figure 1](image1.png)

**Figure 1.** Probability category curves of item 3 ("help me to connect"). On the left is an example of disordered thresholds (response categories not working as intended). On the right is an example of categories working as intended. In the case on the right, the ordered categories represent an increase in latent construct value (recovery orientation) for each response category. More specifically, the x axis (from −3 to + 3) represents increasing amount of recovery-orientation quantified in logits for each response category. The y axis represents the probability of endorsing that item, given a person’s overall total quantification of recovery-orientation. So, as shown on the right, a person who weakly endorses (blue line labeled as “0”) item #3 ["Staff members help me to connect with self-help, peer support, or advocacy groups/programs"] is less likely to score high on the recovery orientation scale when compared to a person who strongly endorses the item (green line labeled as “2”). In the figure on the left, we do not observe this natural ordering of response categories. See the online article for the color version of this figure.

![Figure 2](image2.png)

**Figure 2.** Item map showing a person’s expected score to each item as a function of the measure of recovery-orientation. The items are listed in terms of less (item 2) to more (item 9) recovery orientation. The x axis represents the theoretical continuum of the latent construct (less to more recovery orientation). The y axis lists the items included in the final analysis of 10 items. In this case, the 5 item response categories were collapsed into three response categories (with the middle three response categories collapsed together). The figure depicts response categories that are ordered and working as intended, suggesting a three category response option may be more favorable for this sample under study. See the online article for the color version of this figure.
inherent areas of improvement in the scale’s structure that would not have been picked up by traditional analysis methods. Specifically, information from the Rasch analysis provided important information about the scale’s poor response category structure, misfit of 2 items, and notable poor targeting of items to the sample (specifically at the higher end of the recovery-orientation spectrum as shown in Figure 3). Regarding the RSA-B’s poor response category structure, we found that that a five-category response format for each item in the RSA-B did not work as intended (see example in Figure 1). Specifically, the results highlight that participants were unable to .....

Table 2
Measures of Fit and Location (SE) of RSA-B Items

<table>
<thead>
<tr>
<th>No.</th>
<th>RSA-B items</th>
<th>Location</th>
<th>SE</th>
<th>FitResid</th>
<th>χ²</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Staff believe that I can recover and believe that I can make my own treatment and life choices.</td>
<td>-0.394</td>
<td>0.055</td>
<td>1.478</td>
<td>13.230</td>
<td>0.152</td>
</tr>
<tr>
<td>12</td>
<td>Staff work to follow my choices and preferences.</td>
<td>-0.219</td>
<td>0.054</td>
<td>-3.129*</td>
<td>33.354</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>5</td>
<td>I can/am can be involved in the evaluation of services and service providers (for example: satisfaction questionnaires, asked to provide feedback about services).</td>
<td>-0.130</td>
<td>0.053</td>
<td>1.308</td>
<td>12.153</td>
<td>0.205</td>
</tr>
<tr>
<td>3</td>
<td>Staff members help me to connect with self-help, peer support, or advocacy groups/programs.</td>
<td>-0.050</td>
<td>0.056</td>
<td>-0.770</td>
<td>6.120</td>
<td>0.727</td>
</tr>
<tr>
<td>4</td>
<td>Staff help me to become involved in programs that are not related to mental health or addiction services (for example: church groups, sports teams, and adult education).</td>
<td>-0.032</td>
<td>0.054</td>
<td>-0.031</td>
<td>5.468</td>
<td>0.791</td>
</tr>
<tr>
<td>7</td>
<td>I can look at my treatment or client records, if I wish.</td>
<td>0.030</td>
<td>0.054</td>
<td>1.254</td>
<td>13.564</td>
<td>0.139</td>
</tr>
<tr>
<td>10</td>
<td>Staff help me to develop career and life goals that go beyond managing my symptoms/keeping my mental health stable.</td>
<td>0.066</td>
<td>0.054</td>
<td>0.393</td>
<td>8.964</td>
<td>0.446</td>
</tr>
<tr>
<td>6</td>
<td>Staff try to involve my significant others (spouses, friends, family members) and other important people in my life (e.g., clergy, neighbors, landlords) in the planning of my services, if this is my preference.</td>
<td>0.136</td>
<td>0.052</td>
<td>-0.923</td>
<td>12.396</td>
<td>0.192</td>
</tr>
<tr>
<td>8</td>
<td>A big part of my work with staff involves helping me to develop my leisure interests and hobbies.</td>
<td>0.277</td>
<td>0.050</td>
<td>-0.520</td>
<td>11.815</td>
<td>0.223</td>
</tr>
<tr>
<td>9</td>
<td>Staff give me a chance to talk about my sexuality and/or spiritual needs and interests.</td>
<td>0.317</td>
<td>0.054</td>
<td>0.965</td>
<td>9.684</td>
<td>0.378</td>
</tr>
</tbody>
</table>

Note. Items are located in order of low (Item 2) to higher recovery orientation (Item 9).

A Collapsing categories and excluding items 1 and 11 because of misfit: fit residual > 2.5. * Mean location score obtained for items. In RUMM 2030, the scale is centered on zero logits, representing the item of average difficulty for the scale. c Residual statistics are the standardized sum of all differences between observed and expected values scored over all persons. An item with a residual statistic less or greater than 2.5 warrants further investigation. d Chi-square statistic compares the difference of observed values with expected values across groups with different ability levels (from less to high recovery orientation) across the latent trait being studied (recovery-orientation).

* Fit statistics higher than expected, suggesting this item is not fit for purpose to measure the construct of interest.

Second, our results from the Rasch analysis suggest that that two items in the RSA-B scale did not fit well with the other 10 items used to capture the construct (item 1: “Staff do not threaten, bribe, or force me to do things that I don’t want to do,” and item 11: “Staff members address my personal life experiences, interests, and needs, and also address my unique culture”). Both items had statistical and graphical evidence for not fitting with the other items. On closer inspection, and revisiting the theoretical model for recovery orientation, we confirmed that both items have components that reflect the concept of interest. However, when looking at the wording of each item, we identified that both misfitting items asked about multiple components of the concept, potentially

Table 3
Summary of Measures of Rasch Model Fit* for RSA-B Items

<table>
<thead>
<tr>
<th>Measure of fit</th>
<th>Basal model</th>
<th>Adjusted model</th>
<th>Subtest model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item fit residual (SD)</td>
<td>0.85 (1.17)</td>
<td>0.02 (1.42)</td>
<td>0.30 (2.65)</td>
</tr>
<tr>
<td>Person fit residual (SD)</td>
<td>-0.17 (1.34)</td>
<td>-0.56 (2.02)</td>
<td>-0.59 (1.66)</td>
</tr>
<tr>
<td>Total item χ²</td>
<td>1082.20</td>
<td>126.75</td>
<td>81.594</td>
</tr>
<tr>
<td>χ² P-value</td>
<td>0.000000</td>
<td>0.006517</td>
<td>0.009</td>
</tr>
<tr>
<td>Person Separation Index</td>
<td>0.72</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td>t test P (CI = 95%)</td>
<td>0.74% (0.70-0.81)</td>
<td>1.15% (1.03-1.23)</td>
<td>1.29% (1.21-1.37)</td>
</tr>
</tbody>
</table>

* A Rasch model is a model that represents the structure which data should exhibit to obtain measurements from the data. * Collapsing categories and excluding items 1 and 11. * Subtest analysis: subtest items 2, 3, and 4 and items 9, 10, and 12. d Bonferroni adjusted Chi-Square 0.005.
resulting in confusion in how best to select a response for each item. For example, item 1 asks about “threatening” and “forcing”, two different ideas that participants may have interpreted differently by the participants. Item 11 has four components included in one item. Other researchers have also found that item wording plays an important role in how well items fit to the Rasch model (Salyers et al., 2013). Future attempts to develop new items for the RSA-B should consider designing simple, clear questions that inquire about one part of a construct under study. This may not only improve the scale’s structure, but may also facilitate how future results are interpreted by clients, researchers and clinicians (Food & Drug Administration, 2009; Frank, Basch, & Selby, 2014; Patrick et al., 2007).

Third, the results from both the traditional and Rasch analysis highlighted poor targeting of the RSA-B scale in its current format. In the Rasch analysis, we were able to clearly describe this measurement picture in more detail. For example, as shown in Figure 3, our results suggest that items do not cover the full recovery-orientation theoretical continuum. Specifically, we can visually note that more items are needed to cover the upper end of the continuum to cover the concept of interest in this context of use. The Rasch analysis also showed that the mean fit for the sample was greater than zero. This indicates that recovery orientation ratings of the sample were higher than expected. When looking at Table 1 and Figure 3, we observe a clear ceiling effect. This suggests that the range of data that can be gathered by the RSA-B, specifically at the high end of the recovery orientation spectrum, may be constrained by inherent limits in the rating scale’s design. This result provides important evidence to support the need for future work to generate new items that capture higher levels of recovery-orientation.

The third objective of this study was to help researchers and clinicians increase their awareness of how modern methods can inform decision making in psychiatric rehabilitation. Ideally, a rating scale can show meaningful change in a clinical trial (Hobart et al., 2013). However, the application of rating scales data in psychiatric rehabilitation has significant application for use in front-line clinical decision making and for tracking individual mental health and wellness outcomes. The results from the Rasch analysis suggest that in its current format, many of the items of the RSA-B fall on the lower end of the recovery orientation theoretical continuum. On closer inspection these items focus on hope and involvement in goal setting and decision making. Future work is needed to understand what items are needed to fill the upper end of the continuum. The recent literature on personal recovery notes that emphasis on higher levels of personal recovery, such as citizenship, are needed to enable individuals to thrive in the community (Slade et al., 2014; Williams et al., 2012). Some have also encouraged considering recovery orientation beyond the mental health system to include society’s rules and laws, allocation of resources, tangible and intangible public goods, and wider community attitudes (Henwood & Whitley, 2013). Our results provide an important starting point to begin an iterative discussion with key stakeholders about how to address the measurement gaps in

![Figure 3. Distribution of RSA-B items of people who use Assertive Community Treatment services in the sample, obtained by converting total raw scores into linear measurements. The x axis represents the recovery orientation measurement continuum from low to high. The top bars (above the x axis) represent the distribution of people in the sample, whereas the bottom bars (below the x axis) represent items. Ideal targeting would depict a range of items and people covering +/-4 logits (representing 99.9% of the person and item continuum) with each distribution centered at zero. In this figure, we note that there are few items (bottom bars) that are covering the people (top bars) at the top end of the continuum. The shorted scale is well targeted for the people from ~1.5 to + 1.5 logits. The dark two way arrows highlight the measurement gaps in the recovery orientation continuum. Because most people report scores at the higher end of the continuum, this figure suggests that more items are needed to capture the higher end of the recovery-orientation construct for this sample. See the online article for the color version of this figure.](image-url)
the RSA-B so that it can be used as a clinically meaningful tool to inform practice and modifications to health care services.

The modern methods used in this study have not only assessed the measurement properties of the RSA-B, but has also provided valuable information to clinicians and researchers about how to further understand the concept of “recovery orientation” itself. We are aware of only one other study that has assessed the psychometric properties of a measure of recovery-orientation using Rasch methods (Salyers et al., 2013). Salyers and colleagues (2013) found statistical support for a 10-item unidimensional scale of provider expectations for recovery. Although our study also found borderline statistical support for a 10-item scale of client expectations for recovery, our graphical evidence highlights that more work is needed to refine the measure so that it can capture the full range of the theoretical construct. In future work, it will be important to consider how the concept of interest lends itself to be quantified implicitly from low to high and map out how a person can possibly move along a theoretical continuum to have more or less perception of their services being recovery-orientated.

Measurement of health outcomes in psychiatric rehabilitation, and mental health in general, can often be difficult. The approach to scale development, modification, and testing warrants due diligence and use of available psychometric methods that take into consideration the ordinal nature of the data, and subjectivity of the concepts under investigation. To enhance the overall quality of measurement in mental health research and clinical practice, use of rigorous psychometric methods for quantification of constructs is required. Our study emphasizes the value of clinicians and researchers to compliment traditional broad level reliability and validity indicators with newer approaches such as Rasch Measurement Theory, when developing and reviewing the properties of rating scales. By highlighting key anomalies of the items in the RSA-B, our results provide initial evidence for future ways to improve this scale. A well-established and practical measure of recovery-orientation will have the potential to guide decisions in clinical practice and research. It may also be critical to inform meaningful communication between clients and service providers about how best to optimize the care received by individuals with serious mental illness.

Study Strengths and Limitations

The main strength of this study is the use Rasch measurement methods to investigate the psychometric properties of the RSA-B, an instrument intended for wide use to evaluate Assertive Community Treatment services. In addition, the study was based on a large sample of people who use ACT services. Our study was not without limitations. First, this analysis was based on 12 items that were selected from factor analytic methods. Future work to shorten the measure should consider complimenting factor analytical methods with a theoretical and conceptually driven rationale, in addition to expertise from individuals with lived experience. Second, despite receiving participation from 65% of the target sample and 35% of the available sample of ACT clients, our sample was not random. Consequently, our results may also not generalize to all individuals who use ACT services or other types of community mental health services. Further, because of our sampling methods, we also recognize that the ceiling effects observed in this study may an alternative explanation for the scale not functioning well at the upper end. Third, in order to minimize participant burden, limited demographic information was collected. As a result, we were unable to test the stability of the items for differences in responses based on demographic variables (i.e., age, sex, diagnosis, ACT team). Finally, as the study design was cross-sectional, we were unable to test the effect of time.

Conclusions

The RSA-B was designed as a short instrument to be used in clinical practice and research to measure recovery orientation. Given the amount of work and other resources invested in the development of recovery-orientated programs and systems of support, a key question is how can we best measure this important outcome? The current study found support for the shortened scale’s internal validity, internal consistency reliability, and unidimensionality. However, targeting limitations and limited evidence to support the summation of items to form a total score suggest that the RSA-B would benefit from further development and testing before being widely used in clinical practice and research. In future work, priority should be given to improving the targeting and the categorization of the items. Specifically, items are needed to capture the top end of the construct, and a shorter response scoring structure is needed. Although the RSA-B suffers from measurement challenges, important information was gleaned from this study that can guide efforts for future scale modification.

References


Appendix

Twelve-Item Version of RSA-B

Please indicate the degree to which you feel the following items reflect the activities, values, and practices of your agency.

<table>
<thead>
<tr>
<th>Item</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ACT staff do not threaten, bribe, or force me to do things that I don’t want to do.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. ACT staff believe that I can recover and believe that I can make my own treatment and life choices.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. ACT staff members help me to connect with self-help, peer support, or advocacy groups/programs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. ACT staff help me to become involved in programs that are not related to mental health or addiction services (for example: church groups, sports teams, and adult education).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. I am/can be involved in the evaluation of ACT services and service providers (for example: satisfaction questionnaires, asked to provide feedback about services).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. ACT staff try to involve my significant others (spouses, friends, family members) and other important people in my life (e.g., clergy, neighbors, landlords) in the planning of my services, if this is my preference.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I can look at my treatment or client records, if I wish.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. A big part of my work with ACT staff involves helping me to develop my leisure interests and hobbies.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Staff give me a chance to talk about my sexuality and/or spiritual needs and interests.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. ACT staff help me to develop career and life goals that go beyond managing my symptoms/keeping my mental health stable.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. ACT staff members address my personal life experiences, interests, and needs, and also address my unique culture.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. ACT staff work to follow my choices and preferences.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Received June 23, 2014
Revision received March 5, 2015
Accepted March 9, 2015