



Overview

The Wraparound Fidelity Assessment System (WFAS) is a multi-method approach to assessing the fidelity to the wraparound process, a team-based care planning and care management model for youth with complex needs and their families. WFAS instruments include: 1) Interviews with multiple stakeholders (the **Wraparound Fidelity Index** or WFI); 2) A **Team Observation Measure** (the TOM); 3) A **Document Review Measure** (the DRM), and 4) An instrument to assess the robustness of system support for wraparound (the **Community Supports for Wraparound Inventory**).

The instruments that comprise the WFAS can be used individually or in combination with one another. Psychometrics of WFAS instruments, as well as user feedback, are frequently collected and examined in order to assess the instruments' reliability, validity, and utility. This poster summarizes results of recent analyses of national data from the WFI, version 4; national data from the TOM, version 2; and an inter-rater reliability study of the TOM.

Study 1: Reliability and Validity of the Team Observation Measure

Study 2: Wraparound Fidelity Index, version 4

Introduction

The Team Observation Measure (TOM) assesses adherence to wraparound principles during the course of a wraparound team meeting. Previous studies have documented good **variability and normality**, **internal consistency** (total alpha = .862), and **discriminant validity** (significant higher TOM scores for communities implementing wraparound with training and coaching vs. without such supports) (Bruns & Sather, 2009; Bruns, Sather, Hyde, & Walker, 2008).

Until recently, there had not yet been an attempt to document the **concurrent validity** of the TOM, such as by associating TOM results with results from the WFI for the same sample of youth or sites. Moreover, a recent **inter-rater reliability** study (Cox, Sather, & Bruns, 2009) found that inter-rater agreement was below 60% for nine of the instrument's 78 indicators and that mean Kappa scores for indicators was only .464, far below the .60 threshold for "substantial" agreement (Landis & Koch, 1977).

Methods

The Current Study

This poster summarizes results from two separate studies: (1) A **concurrent validity study** that assessed the association between TOM scores and WFI scores for N=8 sites assessed during the same time period, and (2) a small **inter-rater reliability study** that compared ratings obtained by two raters completing the TOM independently.

Measure

The TOM consists of 20 items. Each of the 10 principles of wraparound (Bruns et al., 2004) are assessed via 2 items of the TOM. Each item includes 3-5 indicators of high-quality wraparound practice that can be observed during the course of a typical wraparound team meeting. Scores for each indicator can be assigned as 0 (not observed), 1 (observed), or 666 (Not Applicable), using detailed scoring rules provided in a TOM User's Manual (Bruns & Sather, 2009). Item scores are calculated based on the number of indicators rated as having been observed. The original TOM (2006 version, with 78 indicators) was employed in the concurrent reliability study, while the revised version (2009 version, 71 indicators) was used in the inter-rater reliability study.

Procedures

Concurrent validity study. To conduct the concurrent reliability study, we obtained data from N=8 sites that had collected data from both the WFI and TOM. Data were collected by local evaluators who were trained to criteria using manualized procedures. N of youth ranged from 10-39 TOMs and 17-207 WFIs.

Inter-rater reliability study. Two raters (a Research Coordinator and Wraparound Program Manager) completed the measure independently as part of an NIMH-funded study of wraparound in Nevada. Both raters were trained to criteria using the TOM User's Manual. TOMs were completed between October 2009 and February 2010 for N=12 wraparound team meetings for N=12 unique youth.

Results

Concurrent validity study. Figure 1 presents mean overall WFI scores (combining scores for multiple respondents) and mean overall TOM scores for the N=8 sites included in analyses. As shown, TOM site means ranged from 56 – 93 percent of the total possible score (mean = 75.2; SD = 12.3), and WFI site means ranged from 69-84 percent (mean = 76.6; SD = 12.3). A highly significant association was found between TOM and WFI mean site-level scores ($r(8) = .857$; $p < .01$)

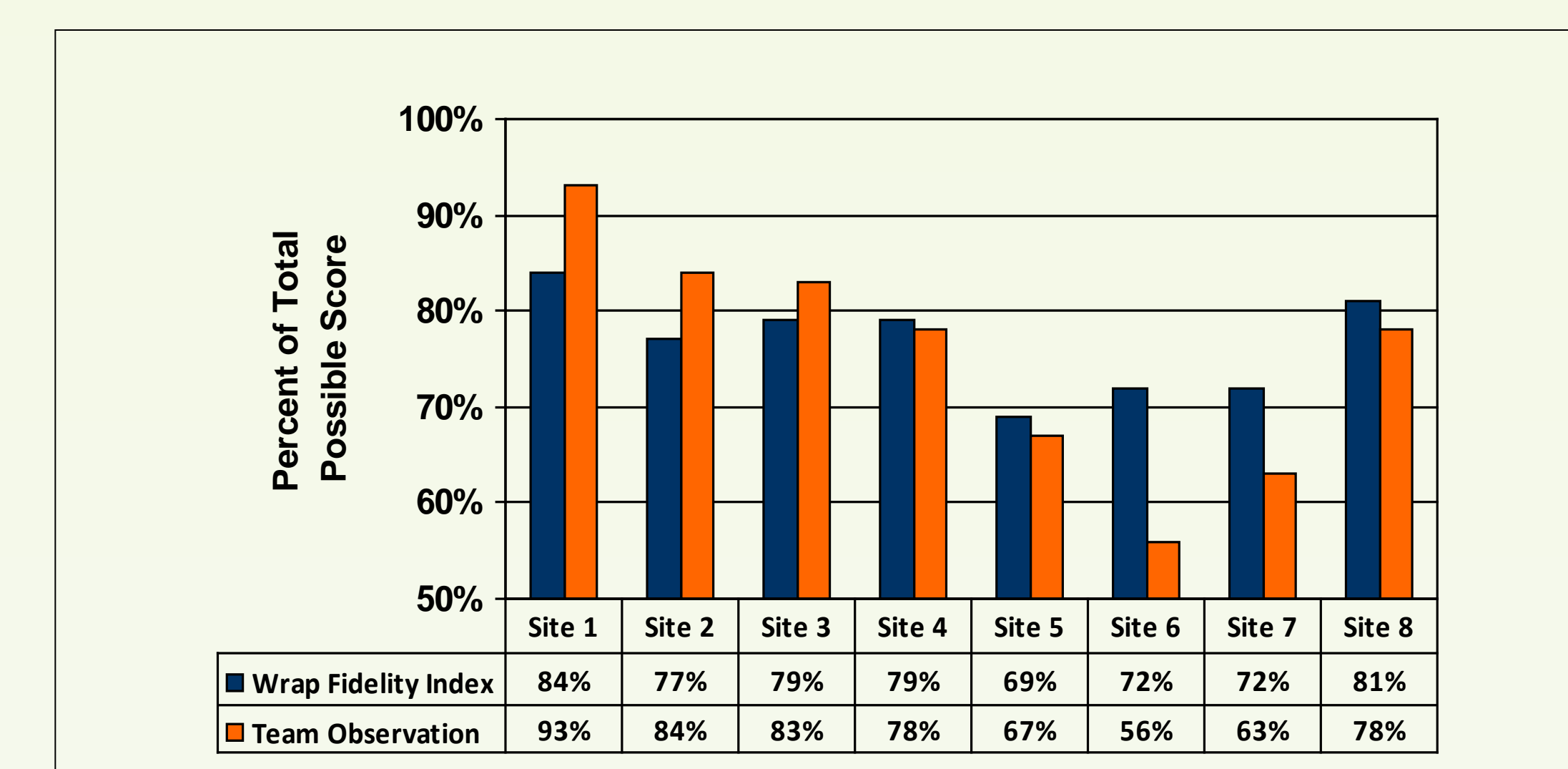


Figure 1. Mean site-level scores on the WFI and TOM for N=8 sites

Inter-rater reliability study. Total TOM scores for the N=12 teams observed were found to range from 49-70 percent of total possible score (mean = 57.7; SD = 7.7). As shown in Figure 2, results of reliability analyses indicate improved inter-rater reliability in the new version of the TOM. Only 2 of 71 TOM indicators (2.8%) were found to have inter-rater agreement poorer than 60%, compared to 9 (11.5%) in the previous study. Moreover, the mean Kappa coefficient (which corrects for agreement that is possible due to chance) was found to be .835 in this study, compared to .464 in the study conducted with the previous scale. Nonetheless, N=7 indicators were found to be difficult for the raters to score reliably, as indicated by Kappa scores < .60.

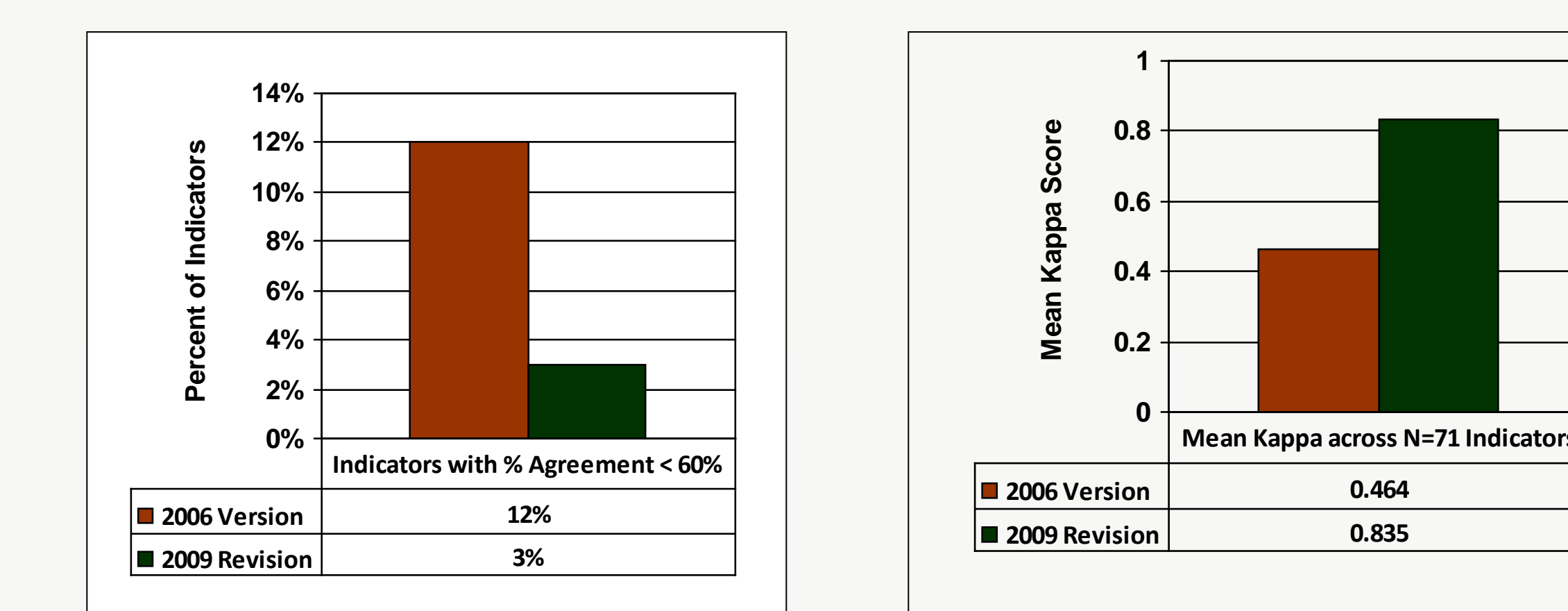


Figure 2. Improvements in Inter-rater Reliability for Team Observation Measure

Discussion

The first study demonstrates encouraging evidence for the validity of the TOM. Even with a small number of sites, the association between mean TOM and WFI scores was robust. The second study suggests that the inter-rater reliability of the TOM has been improved as a result of its revision. Additional reliability testing will also help the research team determine whether indicators found to have poorer reliability in the require revision or deletion from future versions of the TOM.

Introduction

Versions 1-3 of the Wraparound Fidelity Index (WFI) were developed as tools to evaluate the fidelity of the wraparound process to the *principles* of wraparound (Bruns, Suter, Force, Sather, & Leverenz-Brady, 2009). However, these versions of the tool did not examine the *specific activities* of wraparound because there was no national consensus on what these activities were. As one product resulting from the efforts of the National Wraparound Initiative team to build consensus on these activities (Walker, et al., 2004), the WFI version 4.0 (WFI-4) features modifications to the WFI to capture wraparound teams' adherence to these activities. Previous studies examined the WFI versions 1-3 and found adequate reliability, validity, and relationship to child outcomes (Bruns, et al., 2009).

Past studies used a classical test theory (CTT) approach. Item response theory (IRT), which is a contemporary alternative to CTT, has not been conducted with the WFI-4. In modern measurement construction, the goal is to build a test in which all items measure the same construct, but each item adds something to the test which most other items do not measure. IRT models are often superior to CTT in achieving this goal. IRT models rank items based on "difficulty" of "correct" response; in terms of **Wraparound Fidelity**, more "difficult" items are represented by items which only higher-fidelity wraparound teams endorsed.

Methods

Measure

The WFI-4 has versions for wraparound facilitators (WF), caregivers (CG), other team members (TM), and youth (Y). It consists of 40 items for WF, CG, and TM, and 32 items for youth. Items are separated into 4 phases of wraparound: engagement, planning, implementation, and transition. Items are designed to measure activities which fit each of the 10 principles of wraparound. Items are scored as Yes, Sometimes/Somewhat, and No; several items are reverse-scored, and higher scores indicate increased wraparound fidelity.

Procedures

Collaborating sites conducted interviews with caregivers, wraparound facilitators, youth, and other team members ideally at least one month, but possibly up to three months after families began wraparound. Interviewers were trained to criteria via manualized procedures. Data was submitted from 1,478 unique wraparound teams based in 41 different sites, nested within 25 projects, and included data from 1,234 wrap facilitators, 1,006 caregivers, and 221 team members.

Acknowledgments: This work was supported by the National Institute of Mental Health (R41MH077356; R34MH072759) and by the Child, Adolescent and Family Branch of the SAMHSA Center for Mental Health Services.

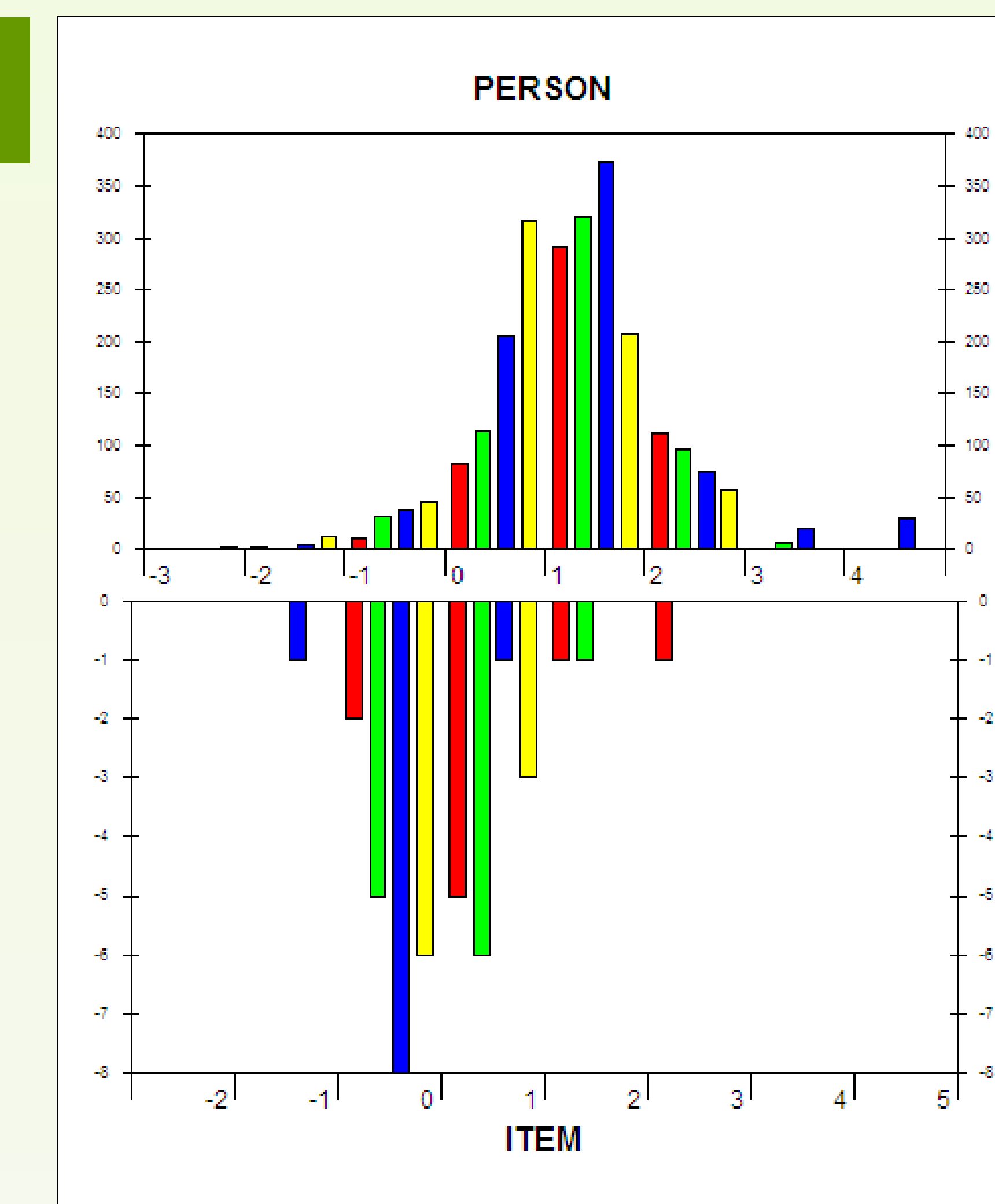


Figure 3. WFI Person-Item "Difficulty" Location

Results

We applied a Rasch partial credit model. Results indicated that the WFI-4 captures a well-defined unidimensional construct, with items ranked in order of "difficulty" in a consistently predictive way, and without items that were inconsistent or surprisingly different than the model would suggest.

Variability accounted for in the data by the model was only 34%, likely a result of a ceiling effect with the measure. This is depicted in Figure 3, which shows the person and item locations scaled by logits from -3 to 5, with increasing values representing increasing wraparound fidelity. The majority of items are located at a level of wraparound fidelity below most respondents, which is related to measurement error. A second aspect of IRT is differential item functioning (DIF), analogous to inter-rater reliability. Three items were considered to have differential functioning:

1.4 "Did family members select the people who would be on their wraparound team?"; 1.5 "Is it difficult to get team members to attend team meetings when they are needed?"; and 3.7 "Does the team come up with new ideas for the wraparound plan whenever the family needs change or something is not working?."

Discussion

These analyses suggest possible modifications to improve the measure. First, several items in the middle level of difficulty could be eliminated from the measure with no loss of value to the scale. Second, additional "difficult" items could be included to more precisely capture the upper range of fidelity. Third, most items behaved in an essentially dichotomous fashion. It may be useful to change the number of response categories from three to two. Fourth, three items displayed differential item functioning, where caregivers and wraparound facilitators at equal levels of WFI-4 scores differed in likelihood of endorsing these items. These items should draw careful consideration and, if bias is considered likely, be modified or removed.

References

- Bruns, E.J., Sather, A.K., Hyde, K., & Walker, J. S. (2008). *The Wraparound Fidelity Assessment System: Measures, Methods, and Data from National Pilot Sites*. Poster Presentation at the 21st Annual Conference "A System of Care for Children's Mental Health: Expanding the Research Base."
- Bruns, E.J., & Sather, A.K. (2009). *Wraparound Fidelity: What Does the National Research Tell Us?* Presentation at the Building on Family Strengths Conference "Research and services in Support of Children and Families."
- Bruns, E.J., & Sather, A.K. (2009). *Team Observation Measure: Manual for training, administration, and scoring of the TOM*. Seattle, WA: Division of Public Behavioral Health & Justice Policy, University of Washington.
- Bruns, E.J., Suter, J., Force, M.M., Sather, A.K., & Leverenz-Brady, K. (2009). *Wraparound Fidelity Index 4.0: Manual for training, administration, and scoring of the WFI 4.0*. Seattle, WA: Division of Public Behavioral Health and Justice Policy, University of Washington.
- Cox, K., Sather, A.K., & Bruns, E.J. (2009). *The Team Observation Measure: Validity, Reliability, and Psychometrics*. Poster Presentation at the Building on Family Strengths Conference "Research and Services in Support of Children and Families."
- Landis, J.F. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Walker, J. S., Bruns, E. J., Adams, J., Miles, P., Osher, T. W., Rast, J., et al. (2004). *Ten principles of the wraparound process*. Portland, OR: National Wraparound Initiative, Research and Training Center on Family Support and Children's Mental Health, Portland State University.