Running head: EVALUATING IFSPs

Using the Rasch Model to Refine the Early Childhood Individualized Family Service Plan Rating Scale

Lee Ann Jung, Shannon O. Sampson and Kelly D. Bradley

University of Kentucky

Correspondence concerning this article should be addressed to Lee Ann Jung, University of Kentucky, ljung@uky.edu or Kelly D. Bradley, University of Kentucky, kdbrad2@uky.edu

Abstract

This study presents evidence regarding the construct validity and internal consistency of the IFSP Rating Scale (McWilliam & Jung, 2001), which was designed to rate individualized family service plans (IFSPs) on 12 indicators of family centered practice. Here, the partial credit Rasch measurement model is employed to investigate the scale's functioning and fit for both person and item diagnostics of 120 IFSPs that were previously analyzed via principal components analysis. Analysis demonstrated scores on the IFSP Rating Scale fit the model well, though additional items could improve the scale's reliability. Implications for using the IFSP Rating Scale in research and practice are discussed.

Evaluating the Construct Validity and Internal Consistency of Early Childhood Individualized Family Service Plans

McWilliam and Jung (2001) constructed the individualized family service plan *(IFSP) Rating Scale* in an effort to create a more standardized evaluation measure of the quality of these plans, which are written for infants and toddlers with disabilities. Studies have been conducted utilizing traditional statistical approaches, such as principal components analysis, to verify the structure of the instrument (Jung & McWilliam, 2005). Even so, such techniques utilize the items as the base of analysis, not giving enough weight to response patterns and theoretical structure. Because IFSPs play a fundamental role in planning early childhood special education services, it is critical to establish construct validity and internal consistency of this rating scale. Given the unidimensional nature of the measure, the Rasch measurement model is the ideal tool to verify the structure of the scale and offer suggestions for improvement.

Background

In 1986, the Individuals with Disabilities Education Act (IDEA) (P.L. 99-457) was amended to include a section regarding Early Intervention (Part C), which includes the provision of developmental and related services to infants and toddlers, birth to 3 years, who have developmental delays or disabilities. Even though states are not required to participate in Part C, all states have elected to do so. By choosing to participate, and in turn, receiving federal funds to support early intervention programs, states must adhere to the regulations set forth by Part C in IDEA. The regulations protect certain rights for families receiving federally funded early intervention, including, but not limited to: developmental assessment for their child, developmental intervention and therapy as appropriate, and an IFSP that guides the process.

The IFSP is a legal document that includes such components as the child's present level of development, family priorities and concerns, outcomes for the child and family, and strategies the intervention team will use. Though federal legislation requires states to include the above components, a national IFSP form does not exist. States may develop their own forms or even allow districts or programs within their states to maintain unique forms. Furthermore, states do not currently use a standard method for evaluating IFSP quality, even though state agencies frequently review IFSPs as a means of evaluating early intervention providers and programs.

An instrument that would guide systematic review and feedback on IFSP quality could prove useful in technical assistance and evaluation efforts. A standard rating scale could also be useful to service coordinators who wish to engage in self-evaluation by objectively examining their own IFSPs. Additionally, it would support IFSP research. The *IFSP Rating Scale* was designed by McWilliam and Jung (2001) for these purposes.

Although estimates of interrater agreement, internal consistency, and component structure in previous studies through classical test theory methods support the notion that quality of IFSPs can be reliably measured using the IFSP Rating Scale (Jung & Baird, 2003; Jung & McWilliam, 2005), application of the Rasch measurement model to the scale can guide refinement of the instrument and provide additional evidence of construct validity. The Rasch model (Rasch, 1980) provides diagnostic information on the quality of a measurement tool. It yields a comprehensive and informative picture of the construct under measurement as well as the respondents' tendencies for that measure. Applying the Rasch model allows researchers to identify where possible misinterpretation occurs in the instrument and which items do not appear to measure the construct of interest. It also produces information about the degree to which each item contributes to the construct. When data fit the Rasch model, it provides estimates for

persons and items that are not dependent upon the particulars of the items used in the questionnaire or of the individuals within the response frame (Wright, 1997; Wright and Masters, 1982). The model provides a mathematically sound alternative to more traditional approaches of rating scale analysis. In this study, the investigators apply the Rating Scale (Andrich 1978a, 1978b, 1979; Wright and Masters 1982) model from the family of Rasch measurement models to the *IFSP Rating Scale* (McWilliam & Jung, 2001) using the 120 IFSPs previously analyzed by Jung and Baird (2003) and Jung and McWilliam (2005).

Methods

This confirmatory study investigates the construct validity and internal consistency of the IFSP Rating Scale (McWilliam & Jung, 2001), designed to rate IFSPs on 12 indicators of family centered practice, via the Rasch measurement model. Using WINSTEPS (3.59.1) software, a rating scale model was applied to test the overall fit of the data to the model. The basic mathematical expression used for constructing measures through responses to the partial credit model is $\log (P_{nij} / P_{ni(j-1)})/B_n - D_i - F_j$ (Andrich, 1978) where P_{nij} represents the probability the person *n* when responding to item *i* would be observed in category *j*; B_n is the attitude of person *n*; D_i is the difficulty of item, and F_j is the "calibration" measure of category j relative to category j-1, the point where categories j-1 and j are equally probable relative to the measure of the item (Linacre, 2005). Here, the scale's functioning and fit for persons and items are assessed on a set of 120 IFSPs that had previously been reviewed using principal components analysis.

Response Frame

Participants were service coordinators in one southeastern state's early intervention system. As a part of a self-study conducted by the state early intervention system, IFSPs were

requested by state administrators from the state's 300 early intervention programs and district early intervention coordinators. Following this request, 135 service coordinators voluntarily provided one IFSP and a non-identifying demographic survey for inclusion in the study. Because service coordinators were instructed to remove all identifying information from these IFSPs to protect child, family, and provider confidentiality, it was not possible to determine which service coordinators responded.

Overall descriptives were produced using the demographic survey responses. Of those responding, service coordinators' experience ranged from one month to over 12 years with a mean of three years. More than half of service coordinators had less than two years of experience. Just over half (53%) reported education-related degrees, and 16% had degrees unrelated to education. Most service coordinators reported having bachelor's (61%) or master's (32%) degrees; 7% reported having associate's degrees. Of the 120 IFSPs submitted, 71% were written by service coordinators who had recently attended a two and one-half day workshop on family-centered practices in natural environments.

Instrumentation

The *IFSP Rating Scale* (McWilliam & Jung, 2001), is a 12-item instrument developed to assess the quality of IFSPs. The instrument is an adaptation of McWilliam's (1993) earlier version and contains items selected based on studies of families' reactions to intervention plans (e.g., Able-Boone, Saridall, Loughry, & Fredrick, 1990; Summers, Dell'Oliver, Turnbull, Benson, Santelli, Campbell, et al., 1990), reflective writings about family-centered intervention plans (e.g., Bailey, Simeonsson, Winton, Huntington, Comfort, Isbell, et al., 1986; Boone, McBride, Swann, Moore, & Drew, 1998; Johnson, McGonigel, & Kaufmann, 1989), curricula for developing family-centered intervention plans (Giangreco, Cloninger, & Iverson, 1993;

McWilliam, 1992), and a review of the literature on family-centered practices in natural environments (Jung, 2001). Some of the items represent the more basic indicators of quality (selecting outcomes that match families' priorities), and others are indicative of the higher level qualities (strategies embedded in routines). Items and descriptions of a 1, 3, and 5 rating for each item are provided in Table 1.

A judge assigns a rating to each item on the *IFSP Rating Scale* according to the quality represented in a given IFSP. Each item is rated on a 5-point scale from 1 (with descriptions of characteristics least consistent with family-centered practice) to 5 (with descriptions of characteristics most consistent with family-centered practice). The rating manual includes directions and examples for determining ratings of 1, 3, or 5. Ratings of 2 and 4 are used when the rater has difficulty determining which rating to assign. For example, if a rater has difficulty determining if an outcome deserves a 1 or 3 on writing, a 2 could be assigned.

Data Analysis

The Rasch Rating Scale model was applied to the data, using WINSTEPS software, version 3.59.1 (Linacre, 2005). The first step in the analysis was to assess the reliability of the measures and to determine the overall fit of the data to the model. Rasch measurement is based on the probabilistic version of the Guttman Scalogram, in which indicators of a construct are arranged from weakest to strongest, and a person who demonstrates a certain indicator is expected to demonstrate all weaker indicators. In order to evaluate how closely the actual data reflect what the measurement system expects, Winsteps calculates fit statistics from individual response residuals. Two chi square ratios are calculated: outfit and infit. The outfit statistics reflect responses that are far off the expectation. In this study, a high outfit value could result from a high quality IFSP with a very low rating on an item that is a weak indicator of quality.

Infit differs from outfit in that each score is weighted so that extreme responses have less influence on the magnitude of the statistic (Bond & Fox, 2001; Wright & Masters, 1982; Wright & Stone, 2004). The summary statistics (Tables 2 and 3) were used to judge the reliability of the measures and the fit of the data to the model. These tables include the mean and standard deviation of the raw scores, the number of times a response was used by a person (Count), the logit conversions of the raw scores (Measure), the error, and the Infit and Outfit values.

The item-IFSP map (Figure 1) was used to look at the hierarchy of the items to determine how the level of quality in an IFSP compares to the amount of quality represented in each item on the IFSP Rating Scale, and to determine how well the items are distributed along the continuum of the variable of IFSP quality. Mean square fit statistics for individual items (Table 4) were used to determine the extent to which each item meets the requirements of the model. A mean square value of 1 is ideal; using Wright and Linacre's (1994, as cited in Bond & Fox, 2001) recommendation, values ranging from 0.6 to 1.4 were considered to have acceptable fit for the items on the IFSP Rating Scale. The model expects that an IFSP receiving a high rating on an item representing a large amount of quality should also receive high ratings on items representing lesser amounts quality; in turn, items representing lower quality should always be assigned higher ratings for any IFSP.

In previous studies, scores for IFSPs' "writing" were treated as one item. Because writing was scored across two parts of the IFSP (present level of development and outcomes), the investigators chose to treat writing as two separate items in this analysis, resulting in a total of 13 items. Because each item on the IFSP Rating Scale was applied multiple times on each IFSP (e.g., a writing score is generated for each outcome), the final item score for the IFSP was recorded as a mean, most often including a decimal. Because Winsteps requires integer data,

before carrying out Rasch analysis, the data were recoded to whole numbers. The IFSP training package indicates that graders are to assign a score of 1, 3, or 5, and to only assign 2s and 4s if an item is between the recommended options. In converting the scores, the researchers determined that items would only receive a 1 or a 5 if these were the assigned scores in the original data. Ratings between 2.5 and 3.5 were coded as a 3. Because a 2 and 4 rating resulted from indecision between a 1 and a 3 or a 3 and a 5, respectively, ratings greater than 1 and less than 2.5 were coded as 2, and ratings greater than 3.5 and less than 5 were coded as 4. This recoding was based upon the original use of the scale, where 1, 3 and 5 served as the anchors. Furthermore, because a single rater assigned scores to all of the IFSPs with a second rater assigning scores to only 12 IFSPs, and because the interrater agreement was high, only the ratings of the principal rater were included in this analysis.

Results

Reliability

The IFSP Rating Scale reliability in this case is the replicability of IFSP ordering one could expect if this sample of IFSPs was rated using another set of items measuring the same construct. Though the IFSP reliability was acceptable at .81, it could be improved through more items targeted to the upper end of the sample (Fox & Jones, 1998, as cited in Bond & Fox, 2001). The item reliability is the "replicability of item placements along the [construct] if these items were given to another sample with comparable ability levels" (Bond & Fox, 2001, p. 32). In other words, if another sample of IFSPs of comparable quality were rated using the *IFSP Rating Scale*, would the estimates indicating the level of quality represented in the item remain stable? Because the reliability estimate is .98, it is likely that these would be consistent estimates for other samples of IFSPs.

Hierarchy of Items

The item map (Figure 1) displays the hierarchy of items rated on the IFSPs. This vertical line is analogous to a ruler. Here, items on the IFSP rating scale (to the right of the vertical line) act as the measurement units to gauge the quality of the IFSPs (listed to the left, with each "X" representing two IFSPs.) Higher quality IFSPs are toward the top of the hierarchy, and lower quality IFSPs fall toward the bottom. IFSP rating scale items that are more indicative of quality, and thus more difficult to rate highly for any IFSP, are listed at the top of the scale. This vertical line is analogous to a ruler. Here, items on the IFSP rating scale (to the right of the vertical line) act as the measurement units to gauge the quality of the IFSPs (listed to the left.) Higher quality IFSPs are toward the top of the hierarchy, and lower quality IFSPs fall toward the bottom. IFSP rating scale items that are more indicative of quality, and thus more difficult to rate highly for any IFSP, are listed at the top of the scale. In other words, a rating of 5 on *location* is easy to obtain, but a high rating on *context appropriateness* is difficult to obtain. The spread of items and IFSPs is good, although the spread of IFSPs extends beyond the items at the upper end of the scale, and the spread of items extends beyond the IFSPs at the lower end. In order to achieve a more accurate rating of IFSPs at the top end of the scale, more scrutinizing items could be added to the rating scale. As noted earlier, this would also increase the reliability of the instrument.

Item Fit

Table 3 displays the fit of the item data to the Rasch model. Using Wright & Linacre's (1994, as cited in Bond & Fox, 2001) recommendations for a rating scale, the mean square range considered acceptable is .6 to 1.4. On the upper end, all items fit with that expectation. On the lower end, Positiveness and Match Outcome are the two that misfit. Fit statistics of less than .6 indicate overfit, which means the item has less variation than the model expects. This is not as

much of a concern, as the validity of actual decisions about individuals based on their measure is rarely threatened by overfit. Overfitting items rarely have any practical consequences (Linacre, 2000).

Figure 2 displays ratings assigned to each item on one IFSP with high mean square scores. The column of numbers represents the expected rating for each item for an IFSP with a measure of .73. Observed ratings are indicated with a period on either side of a number (e.g., .5.). When the observed rating is very unexpected, it is placed between parentheses (e.g., (5). This IFSP received 1s on Specific, Active, and Writing (Outcomes) whereas the model predicted higher ratings of 3 on each of these items. On the other hand, this IFSP received a rating of 5 on the two items indicative of a high-quality IFSP. Five of the IFSPs received high ratings on Context Appropriate and Family's Role even though they were weak in many less difficult areas.

Discussion

The order of items on the IFSP Rating Scale was consistent with the expectations based on the Rasch model. In other words, items that were rated high on most IFSPs were those that we would expect would be easiest based on what we know about the state of early intervention practice and the political context. Specifically, items that were rated high on most IFSPs were those that either had to do with basic skills on selecting outcomes or were required by law. Basic-skill outcomes should address functional skills, describe a child's strengths, and be based on family priorities. Legally required features include the fact that services should happen in a "natural environment" or in a location where children with out disabilities can be found (e.g., childcare, home). For example, in the state where sample selection took place, administrators had recently instituted a policy of reimbursing intervention only when it was carried out in a natural environment, unless outcomes could not be achieved in the natural environment. Likely

because of this, the item *location* was the item that was most often scored high on these IFSPs. The items that resulted in placement in the middle of the difficulty scale were those that measured whether outcomes were worded in an understandable way (e.g., writing in outcomes), whether the active voice was used, and whether the outcome was measurable (specific).

On the other hand, the items associated with the lowest ratings, *family's role* and *context appropriate,* required that the IFSP team not only select appropriate outcomes and word them well, but also demonstrate current best practice for working with families in the field of early childhood special education. Specifically, high ratings on these items suggest they understand that children learn best in routines and families (a) have the power to make change, (b) can implement strategies, and (c) can implement these strategies most easily in the context of routines.

Even though *family's role* and *context appropriate* were the items that were the most difficult on the scale, there were instances of unexpected, high scores on these items. That is, there were IFSPs that scored higher on these two hardest items but scored lower on the easier items. Figure 2 presents the expected and observed scores on one such misfitting IFSP. One explanation for this is that these higher two items, though they are still part of the quality construct, are indicative of a different skill set that the other items. Specifically, to score well on the easier items requires that the service coordinator knew how to select appropriate outcomes and strategies, and word these well. To score well on the harder two items requires an understanding of current best practice in family-centered implementation of early intervention. Although it is unlikely, it is possible to have grasped best practice for service implementation without having mastered the nuances of completing the IFSP form. In other words, even though each of these is part of a quality IFSP, and the items that were indicated by this analysis to be

harder are indicative of higher skills, these higher skills are not purely dependent on the mastery of the lower skills.

Limitations

Although overall reliability was acceptable, the instrument does not have all the items it needs for maximum clarity and utility. In other words, when Rasch analysis placed items and IFSPs along a "ruler," it indicated where the IFSPs fell (based on their quality) in comparison to where the items fell (based on the level of quality they represented). Just as a ruler that is marked with meters would be of little use in measuring the length of a small insect, when a number of IFSPs has a quality measure that does not correspond to any item's measure, the instrument is considered to be limited in its ability to accurately measure those persons. A well designed instrument has a distribution of items that is equivalent to the distribution of IFSPs (Bradley & Sampson, in press).

This analysis demonstrated, as seen in Figure 1, that there are IFSPs in this sample that have a quality measure that does not correspond to an item. The reliability of the rating scale would be improved by adding additional items that fit in these gaps. Some possible items to consider adding could include ones that determine whether a primary service provider model was used and whether concerns and priorities seem to be based on routines, providing evidence on whether providers interviewed families about their routines early in the IFSP process. Because the instrument does include items that measure the most important areas of quality in IFSPs based on the literature, increasing the number of items may not improve the practical utility of the instrument. Furthermore, because the instrument in its current form takes approximately 45 minutes to complete, additional items could increase the time needed to administer beyond the time providers are willing to spend.

Conclusion

Although all states receiving federal funding for early intervention try to ensure that IFSPs are in place for families of children receiving services and they include federal and state requirements, much less attention has been paid to the quality of these plans relative to best practices. Despite the fact that all children who receive federally supported early intervention services have an IFSP, there is no instrument or standardized method that is widely used to evaluate the quality of these plans. Technical assistance providers in many states review and evaluate early intervention services based in part on the quality of IFSPs. The *IFSP Rating Scale* would be appropriate for such personnel who need a measure of IFSP quality that can guide their feedback and support.

Many providers in early intervention are eager for guidance and suggestions on how to write IFSPs. The *IFSP Rating Scale* provides specific criteria and examples of elements of good practice. The scale could prove very useful to those who want to evaluate their own IFSPs within the context of individual families. Evidence from this and previous studies also indicates that the IFSP Rating Scale is appropriate for research on IFSP quality. National or multi-state studies could provide generalizable information on the quality of IFSPs that could inform policy and professional development.

The IFSP is the one document that is common to *all* families in early intervention. It is written with and for families very early in the process. It tells families implicitly and explicitly what to expect from the professionals from whom they will receive support. Because of this, the quality of these documents deserves attention. The IFSP Rating Scale may be one tool to influence thoughtful preparation and evaluation at individual, systems, and research levels.

References

Able-Boone, H., Sanridall, S. R., Loughry, A., & Frederick, L. L. (1990). An informed, family-centered approach to Public Law 99-457: Parental views. *Topics in Early Childhood Special Education, 10,* 100-111.

Andrich, D. (1978a). Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement, 38,* 665-680.

Andrich, D. (1978b). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2,* 581-594.

Bailey, D. B., Simeonsson, R. J., Winton, P. J., Huntington, G. S., Comfort, M., Isbell, P., O'Donnell, K. J., & Helm, J. M. et al.. (1986). Family-focused intervention: A functional model for planning, implementing, and evaluating individualized family services in early intervention. *Journal of the Division for Early Childhood, 10,* 156-171.

Bailey, D. B., Winton, P., Rouse, L., & Turnbull, A. (1990). Family goals in infant intervention: Analysis and issues. Journal of Early Intervention, 14(1), 15-26.

Bond, T., & Fox, C. (2001). *Applying the Rasch model: Fundamental measurement in the Human sciences.* New Jersey: Lawrence Erlbaum Associates.

Boone, H. A., McBride, S. L., Swann, D., Moore, S., & Drew, B. S. (1998). IFSP practices in two states: Implications for practice. *Infants and Young Children, 10(*4), 36-45.

Boone, H. A., Moore, S. M., & Coulter, D. K. (1995). Achieving family-centered practiced in early intervention. *Infant Toddler Intervention*, 5, 395-404.

Bradley, K. D. and Sampson, S. O. (2006). Chapter 2. Utilizing the Rasch Model in the Construction of Science Assessments: The Process of Measurement, Feedback, Reflection and Change In X. Liu and W. Boone (eds.), Applications of Rasch

measurement in science education (pp. 23-44). Maple Grove, MN: JAM Press.

Bruder, M. B. (2000). The individualized family service plan (IFSP). (ERIC EC Digest #E605).

Bruder, M. B., Staff, I., & McMurrer-Kaminer, E. (1997). Toddlers receiving early intervention in childcare centers: A description of a service delivery system. *Topics in Early Childhood Special Education, 17,* 185-208.

Dunst, C. J. (1999). Placing parent education in conceptual and empirical context. *Topics in Early Childhood Special Education 19,* 141-147.

Fox, C. M., & Jones, J.A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology, 45*(1), 30-45.

Giangreco, M. F., Cloninger, C. J., & Iverson, V. S. (1993). *Choosing options and accommodations for children: A guide to planning inclusive education.* Baltimore: Brookes.

Individuals with Disabilities Education Act, Amendments of 1997, PL 105-17, United States Code, 20, sections1400 et seq.

Johnson, B. H., McGonigel, M. J., and Kaufmann, R. K. (1989). *Guidelines and recommended practices for the individualized family service plan*. Chapel Hill, NC, & Washington, DC: NEC*TAS and ACCH.

Individuals with Disabilities Education Act, Amendments of 1986, PL 99-457, United States Code, 20, sections1400 et seq.

Jung, L. A., & McWilliam, R. A. (2005). Reliability and validity of scores on the IFSP rating scale. *Journal of Early Intervention, 27,* 125-136.

Jung, L. A., & Baird, S. M. (2003). Effects of service coordinator variables on individualized family service plans. *Journal of Early Intervention, 25,* 206-218.

Linacre, M. (2005)

Linacre, J. M.(2000). Redundant Items, Overfit and Measure Bias *Rasch Measurement Transactions. 14*(3), *755* Retrieved from http://www.rasch.org/rmt/rmt143a.htm.

McWilliam, R. A., & Jung, L. A. (2001). *IFSP Rating Scale*. Unpublished instrument.

McWilliam, R., Ferguson, A., Harbin, G., Porter, D. M., & Vaderviere, P. (1998). The family-centeredness of individualized family services plans. *Topics in Early Childhood Special Education, 18,* 69-82.

McWilliam, R. (1993). *The IFSP Family-Centeredness Rating Scale.* Frank Porter Graham Child Development Center, University of North Carolina at Chapel Hill.

McWilliam, R. A. (1992*). Family-centered intervention planning: A routines-based approach.* Tucson, AZ: Communication Skill Builders.

Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests (revised and expanded ed.). Chicago: The University of Chicago Press. Sandall, S., McLean, M. E., & Smith, B. J. (2000). DEC Recommended practices in early intervention/early childhood special education. Longmont, CO: Sopris West.

Summers, J. A., Dell'Oliver, C., Turnbull, A. P., Benson, H. A., Santelli, E., Campbell, M., & Siegel-Causey, E. (1990). Examining the individualized family service plan process: What are family and practitioner preferences? *Topics in Early Childhood Special Education, 10,* 78-99.

Wright, B. (1997). Fundamental measurement for outcome evaluation [Electronic version]. *Physical Medicine and Rehabilitation: State Of The Art Reviews, 11*(2), 261-288. Retrieved July 8, 2005, from http://www.rasch.org/memo66.htm

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Wright, B.D. and Stone, M.H. (2004). *Making Measures*. Chicago: Phaneron Press

Table 1

Item and Rating Descriptions

Indicator	Rating Methods					
	1	2	3	4	5	Other
Writing	Technical jargon or obscure, complicated language		No technical terms or professional jargon, but statement is unclear		Language is clear and contains no technical jargon	
Active	Passive voice		Active voice/passive outcome		Active voice	
Positiveness	Emphasizes deficits		More information about needs than strengths		At least as many words to describe strengths as needs	
Necessity	Not likely necessary for development		Not clear whether the outcome was necessary		Necessary for development or to support a family routine	
Specificity	Outcome is vague		Outcome not defined and not easily measurable		Sufficiently detailed and measurable	
Context Appropriateness	Unlikely to be incorporated in routines		Restricted the context to one specific activity		Routines suggested	
Match Outcome	Procedures did not seem related to the outcome		Procedures indirectly addressed the outcome		Procedures directly supported the outcome	
Location	Segregated setting		Mixed model		Natural environments only	
Family's Role	Carried out exclusively by professionals		Implied participation from the natural caregivers		Natural caregiver listed as person responsible	
Present Level Functionality	Primarily consisted of tasks from the instrument		Instrument tasks as well as functional information		Included only functional information	
Concerns with Outcomes					Percentage of priorities or concerns addressed by at least one outcome.	
Outcomes with Concerns					Percentage of outcomes that have a corresponding concern or priority	

Table 2

Summary Statistics by IFSP

	Raw Score	Count	Measure	Model Error	Infit MSNQ	Outfit MSNQ
Mean	49.7	14.0	.67	.30	.98	1.00
SD	11.9	2.2	.78	.09	.56	.97
Max	72.0	15.0	2.75	.69	3.09	9.90
Min	18.0	8.0	-.57	.25	.22	.26

	Adj SD	.70	Separation	2.09	IFSP Reliability	.81
	Adj SD	.72	Separation	2.30	IFSP Reliability	.84

SE of the IFSP Mean =.07

Valid Responses: 93.6%

IFSP Raw Score-to-measure Correlation .60 (Approximate due to missing data)

Cronbach Alpha IFSP Raw Score Reliability .91 (Approximate due to missing data)

Table 3

Summary Statistics by Item

	Raw Score	Count	Measure	Model Error	Infit MSNQ	Outfit MSNQ
Mean	391.2	113.2	.00	.10	1.03	.99
SD	84.5	8.5	.75	.02	.34	.25
Max	535.0	120.0	1.24	.15	1.48	1.37
Min	256.0	100.0	-1.18	.08	.51	.53

	Adj SD	.74	Separation	6.76	IFSP Reliability	.98
	Adj SD	1.00	Separation	7.42	IFSP Reliability	.98

SE of the IFSP Mean =.21

Table 4

Rasch Fit Statistics for the IFSP Rating Scale Items

Entry	Raw			Model	Infit		Outfit		Category
Number	Score	Count	Measure	SE	MNSQ	ZSTD	MNSQ	ZSTD	
11	454	100	-1.18	.15	1.18	2.1	1.37	1.5	Location
15	535	120	-1.02	.12	1.42	2.1	1.17	.8	Outcomes with Concerns
12	301	119	.87	.09	1.4	2.9	1.17	1.2	Family's Role
5	315	118	.72	.08	1.3	2.4	1.27	1.9	Active
14	520	120	-.82	.11	1.26	1.5	1.09	.5	Outcomes with Concerns
9	256	119	1.24	.09	1.21	1.5	1.23	1.3	Context
4	345	120	.58	.08	1.19	1.7	1.08	.6	Writing Outcomes
13	355	102	.06	.09	1.06	.5	1.13	.9	Necessity
8	317	120	.77	.08	1.02	.2	.88	-.9	Specific
1	433	102	-.69	.11	.57	-2.9	.81	-.9	Present
7	510	120	-.70	.11	.76	-1.6	.65	-2.1	Necessity
3	350	102	.10	.09	.63	-3.5	.75	-1.9	Writing Present
6	391	103	-.20	.09	.57	-3.7	.70	-2.0	Positiveness
10	395	120	.24	.08	.51	-5.3	.53	-4.3	Match Outcome
Mean	391.2	113.2	.00	.10	1.03	-.2	.99	-.2	
SD	84.5	8.5	.75	.02	.34	2.6	.25	1.7	

Figure 1. Item map showing the distribution of IFSPs and IFSP rating scale items.

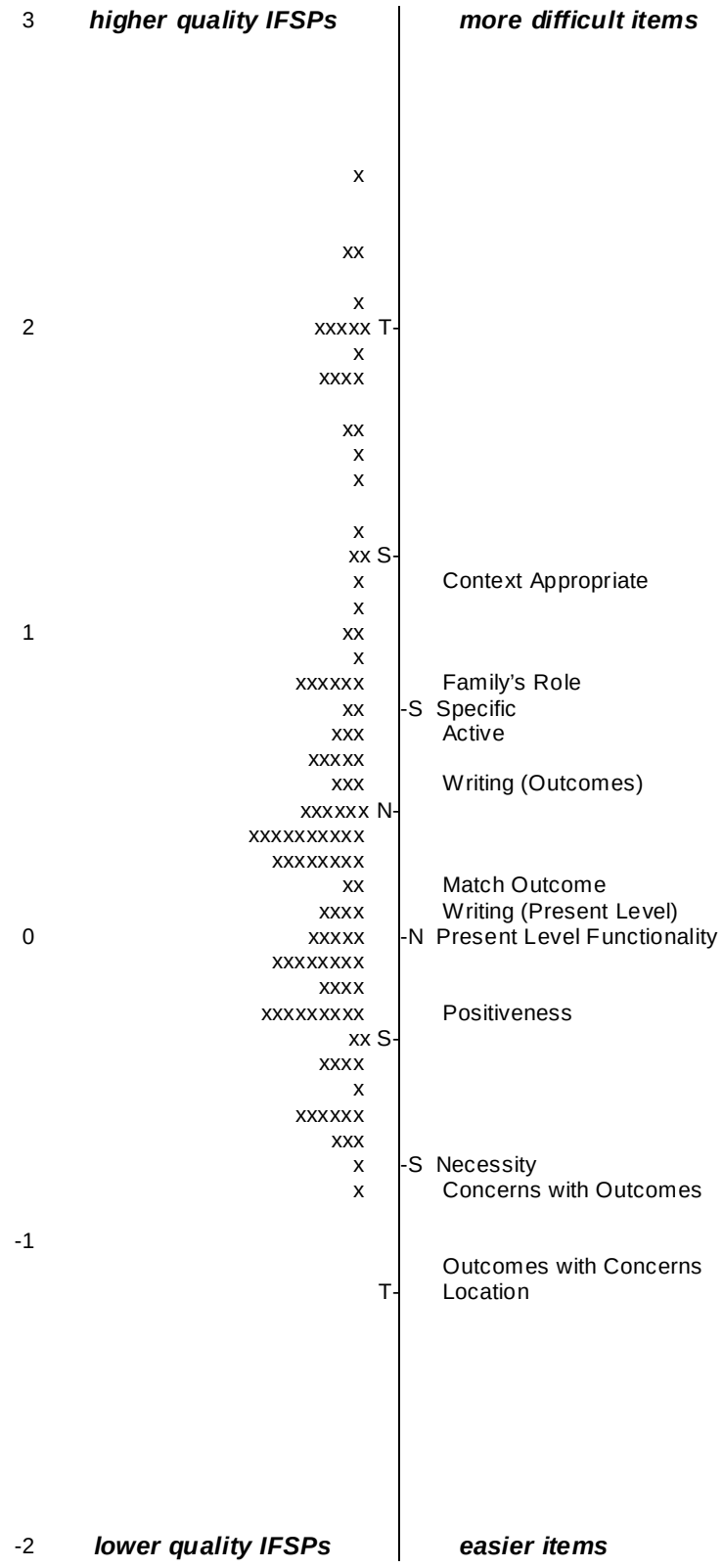

Figure 2. Example of one misfitting IFSP.

-3	-2	-1	0	1	2	Item
			2			Context Appropriate
(5)						
			2			FamilyÕs Role
(5)						
		.1.	2			Specific
		.1.	2			Active
		.1.	3			Writing (Outcomes)
			.3.			Match Outcome
			3	.4.		Writing (Present Level)
			.3.			Present Level Functionality
			4		.5.	Positiveness
		.3.	4			Necessity
		.3.	4			Concerns with Outcomes
		.3.	4			Outcomes with Concerns
			.5.			Location
-3	-2	-1	0	1	2	

= Expected Score
.#. = Observed Score
(#) = Unexpected Score