

Construction and Validation of Two Parent-Report Scales for the Evaluation of Early Intervention Programs

August 2011

William P. Fisher, Jr., University of California, Berkeley
Batya Elbaum, University of Miami
W. Alan Coulter, Louisiana State University Health Sciences Center



Summary

The State Performance Plan (SPP) developed under the 2004 reauthorization of the Individuals with Disabilities Education Act (IDEA 2004, Public Law 108-446) requires states to collect data and report on three key outcomes of early intervention (EI) services for participating families. The Impact of EI Services on the Family Scale (IFS) and the Family Centered Services Scale (FCSS) were developed to provide states with a means to address this new reporting requirement and to collect additional data that would inform program improvement efforts. Items suggested by stakeholder groups were piloted with a nationally representative sample of parents of children with developmental delays or disabilities ages birth to three participating in EI services in eight states. The 28-item IFS had measurement reliabilities ranging from .93-.96 in a sample of 1,750; measurement reliabilities for the 135-item FCSS ranged from .94-.97 in a sample of 1,755 respondents. A 29-item version of the FCSS had measurement reliabilities ranging from .87-.92. Using data from the pilot study, stakeholders established a recommended performance standard set at a meaningful point in the IFS item hierarchy, for each of the three established outcome areas.

Correspondence

Correspondence concerning this article should be addressed to: William P. Fisher, Jr., Graduate School of Education, 1501 Tolman Hall, University of California, Berkeley, CA 94720-1670. E-mail: wfisher@berkeley.edu

Acknowledgements

This research was supported in part by U.S. Department of Education, Office of Special Education Programs grant award #H373Y070002 to Westat.

The authors also thank Larry Wexler, Jack Stenner, and Michael Everett for their support of this work.

Introduction

Recognition of the importance of families to the healthy development of their children is one of the foundations of early intervention services provided to infants and toddlers with developmental delays or disabilities. Within a family-centered perspective, the best interests of the child are seen as being served not only by providing specific therapies for the child but also by strengthening the ability of the family to promote the child's positive development. Supports and services for the family and child must be provided in a way that takes into consideration the family's routines, preferences, and cultural practices.

The new federal accountability system for early intervention services, developed under the re-authorized Individuals with Disabilities Education Act (Public Law 108-446; IDEA 2004), supports the principle of family-centered services by including, among the 14 performance indicators on which states must collect data and report every year, an indicator (Indicator C-4) related to family outcomes of early intervention. Specifically, states must report the percentage of families participating in early intervention services who report that early intervention helped them a) know their rights, b) effectively communicate their children's needs, and c) help their children develop and learn (Regional Resource Center Program, 2008).

In 2002, the National Center for Special Education Accountability Monitoring (NCSEAM), a technical assistance center funded by the U.S. Department of Education, Office of Special Education Programs (OSEP) took on responsibility for developing and validating a family survey that would function as a rigorous measurement tool for states' use in addressing Indicator C-4.

In order to ensure that the measurement tool would meet the most stringent measurement standards, including demonstrated scale invariance across different population subgroups and high measurement reliability ($>.90$), NCSEAM adopted the Rasch measurement framework (Bond & Fox, 2007; Fisher & Wright, 1994; Wright & Stone, 1999) to guide evaluation of item fit, construct validation, and scale refinement. NCSEAM also engaged a national group of stakeholders in establishing recommended standards for reporting purposes, i.e., measures representing the minimum amount of impact on the family that would represent adequate performance in each outcome area. The percent of families with measures at or above the recommended standard in each outcome area is reported in each state's Annual Performance Report.

Method

Participants

Eight state lead agencies for early intervention agreed to participate in the national field study. Participants (families who returned a survey form) consisted of 1,816 families of a child birth to age three receiving early intervention services. The distributions of participants by state of residence, race/ethnicity, child's special needs, child's age, survey form, language in which the survey was completed, and administration mode (self-administered or facilitated) are displayed in Table 1. For comparison purposes, the rightmost column of Table 1 displays the distribution of children ages birth to three served under IDEA by race/ethnicity for the year in which the pilot study was conducted (Data Accountability Center, 2011). As seen in Table 1, the distribution of race/ethnicity in the survey sample is very close to that of the distribution in the population served by state early intervention programs.

Procedure

Item development. In spring of 2003, the authors conducted stakeholder input sessions in six states: California, Florida, Kentucky, New Hampshire, New Mexico, and Mississippi. Between 25 and 42 stakeholders participated in each of the sessions. These individuals included family members of children with disabilities, parent advocates, early intervention providers, state lead agency personnel, and community representatives.

At each meeting, participants were invited to consider multiple aspects of families' experiences with services provided under IDEA and generate statements related to these experiences. Topics included the ways in which early intervention programs and providers meet families' needs, the benefits families experience as a result of their participation in services, and the outcomes of services for infants and toddlers. A total of 170 items related to families' experiences with early intervention services were selected for piloting. Twenty-eight of the items to be piloted addressed the impact of early intervention services on families, 135 addressed the quality of early intervention service delivery, and seven addressed outcomes for children. The present report describes the scales that were developed to address the first two of these areas, namely, the impact of early intervention services on family outcomes and the quality of family-centered early intervention services. All of the items were translated into Spanish by a professional translator, and then back-translated into English by a second translator to ensure equivalence of meaning across the two languages. Several states that planned to administer the surveys online

provided translations in other languages with significant representation in their population.

Pilot study. Approval was obtained from the Louisiana State University Institutional Review Board to conduct a national pilot study of the items generated by stakeholder groups. State lead agencies for EI programs in eight states (California, Florida, Georgia, Iowa, Louisiana, Massachusetts, New Jersey, and New Mexico) agreed to solicit the participation of families in their states who were participating in early intervention services. A customized sampling plan was designed for each state so that the distribution of race/ethnicity in the sample would be close to that of the U.S. population of families served in IDEA-supported EI programs. The sampling plan was also designed to yield approximately equal numbers of children across the three age groups of birth to 1, 1 to 2, and 2 to 3.

NCSEAM arranged for optically scannable forms to be printed and sent to participating states. The way the survey was distributed and administered varied by state. In most cases, the survey form, plus a cover letter explaining the purpose of the survey, was mailed to families with a postage prepaid envelope for return of the completed survey. The survey and cover letter were sent in both English and Spanish. Several states organized group sessions at which a facilitator was available to assist families in completing the survey. Several states offered families the option of responding to an online version of the survey. No personally identifying information was obtained from any of the participating families.

Surveys were scanned by a subcontracting data processing firm which aggregated the records into an SPSS file provided to the authors. Data analyses were conducted using SPSS (2003) and Winsteps (Linacre, 2010).

Instrument

The items related to the impact of early intervention services on the family and a subset ($n = 29$) of those related to early intervention service delivery are presented in Tables 2 and 3, respectively. An identical response scale was used for all items. The response choices were: Very Strongly Disagree, Strongly Disagree, Disagree, Agree, Strongly Agree, and Very Strongly Agree. Respondents were told that they could skip any item they felt did not apply to them or their child.

The readability of the items was assessed using the Lexile Framework for Reading (Stenner, Burdick, Sanford, & Burdick, 2006). Reading difficulties for the items ranged from 850 to 1020 Lexiles, which corresponds to the range of reading difficulty found in books written for sixth-graders.

Two alternate survey forms were developed so that responding families would not have to respond to all 170 items. The 28 items addressing family

outcomes were included in each of two alternate forms; the other 135 items were divided across two forms, with about one-third of the items included on both forms.

Results

Scaling of Items

Item responses from the 1,816 records containing parent responses to the NCSEAM pilot survey were analyzed through the Winsteps software program (Linacre, 2010) using a rating scale model for ordered categories (Andrich, 1978, 2009; Luo, 2005; Wright & Masters, 1982). The item responses were coded from 1 (Very Strongly Disagree) to 6 (Very Strongly Agree). The scale was calibrated to measure along a 0-1000 range, with each original logit corresponding with a rescaled 58.91 units.

IFS. A total of 66 responding families did not provide any rating scale response data on the items related to the IFS, yielding 1,750 measures. Ceiling and floor effects were minimal to small (less than 1% to 11%). The IFS data were about 94% complete.

FCSS. A total of 61 responding families did not provide any rating scale response data on the items related to the FCSS, resulting in estimation of 1,755 measures. Ceiling and floor effects were again minimal to small (less than 1% to 11%). The FCSS data were about 68% complete, due to the distribution of the items across two forms.

Scale Optimization

Examination of the response percentages in each rating category (some of which had fewer than 1% of all responses), of the response probability curves for the original six response categories, and of evidence relating to several other criteria indicated that optimizing the response categories (Andrich, 1996, 2009; Linacre, 1999, 2002, 2004; Luo, 2005) would increase the validity and interpretability of the measures. For meaningful inferences to be made from quantitative measures, observations must be scored so that higher ratings are consistently associated with higher measures, and so that each rating category represents a minimum amount of additional information. Though most approaches to survey data take that association and the increment in information provided for granted, analyses may fail to show support for the implied hypothesis. Close study of disordered category transition thresholds can suggest rescorings likely to result in a closer correspondence of measures with the response probabilities for each category.

IFS. The original six-category response structure for the IFS (Figure 1) was disordered, with the transition threshold between categories 1 and 3 located lower on the scale than that between 1 and 2. The six categories were accordingly reduced to five by collapsing the two lowest categories of disagreement (Very Strongly Disagree and Strongly Disagree) into one category. The probability curves for the optimized, five-category response scale are displayed in Figure 2.

FCSS. For the FCSS, the original six-category response structure (Figure 3) was reduced to a three-category response structure by collapsing all three categories of disagreement into one category, and by combining the two highest categories of agreement (Strongly Agree and Very Strongly Agree) into a single rating. In contrast with the IFS, two of the FCSS categories' transition thresholds were spaced too near the previous categories' thresholds to support the inference of a significant associated increase in the measures. The response probabilities for categories 4, 5, and 6, for instance, are all about .30 in the narrow range where they meet in the right side of Figure 3, meaning that respondents with measures that differ by less than an error are almost equally likely to agree, strongly agree, or very strongly agree. Combining the Strongly Agree and the Very Strongly Agree responses into a single category creates a clearly interpretable transition that stands for a substantive and consistent increase in agreement. The probability curves for the optimized, three-category FCSS response scale are displayed in Figure 4.

Evaluation of Scale Invariance

IFS. Summary statistics from the analyses of the two scales are provided in Table 4. The point-measure correlations (Stenner, 1995) for the IFS items were all positive and ranged from .70 to .82. The item mean square infit statistics ranged from 0.70 to 1.56. Omitting the two items in the IFS with infit statistics over 1.30 had no appreciable effect on the measures beyond a small increase in measurement error. Given the sample size and numbers of rating scale categories, these results indicate data consistency of the quality required for meaningful and useful interpretations of the measures (Linacre, 2003b; R. Smith, 1996b; E. Smith, 2002).

Separate calibrations of the items were conducted on demographically distinct subsamples of respondents defined by state of residence, age of child, family ethnicity, child's disability, language (English vs. Spanish), paper vs. web administration, and whether the respondent read the survey independently or had it read to her or him. Overall, correlations were greater than .78, with the vast majority over .90; disattenuated for error (Muchinsky, 1996; Schumacker, 1996; Wright, 1991), they were all virtually 1.00. The correlations of the by-state IFS calibrations ranged from .82 to .94; 25 of the 28 were greater than .85.

The three correlations between separately estimated IFS calibrations for three different subsamples based in child ages (0-1, 1-2, and 2-3) ranged from .93 to .97. The correlation of the IFS calibrations produced by data from families who filled out the form independently with those produced by families who had the survey read to them was .89. The IFS calibrations correlated .98 when estimated from the 416 persons completing the survey online vs. the 1,334 who used paper forms.

Figure 5 is a scatter plot of the IFS calibrations produced from separate analyses of responses from the families of children age 0-1 versus the families of children ages 1-2. This is a typical plot, with a correlation of .96. Figure 6 is a scatter plot of the IFS calibrations produced from separate analyses of responses from families who used the English-language forms and those who used the Spanish-language forms. The correlation for these calibrations, $r = .78$, was the lowest found in this data set. Disattenuated for the lower reliability (.71) and higher error associated with the smaller Spanish-language sample size (177), the correlation is 1.00.

FCSS. Point-measure correlations were all positive for the FCSS items and ranged from .58 to .74; these are lower for the FCSS than they were for the IFS due to the reduced number of rating scale points on the FCSS. The item mean square infit statistics ranged from 0.63 to 2.31. All of the items with high mean square values fell in the upper end of the scale. Omitting the two items in the FCSS with infit statistics over 1.91, or the eight with infits over 1.76, had no appreciable effects on the measures beyond small increases in measurement error.

Though the availability of a large bank of calibrated items will eventually prove useful for online or computerized adaptive administrations of the FCSS, its practical use in a paper-and-pencil format required a reduction in the number of items. Items were selected on the basis of their overall salience, the consistency of their responses, and so as to represent the full range of the scale. Expectations for the measurement separation reliability likely to be obtained using the shorter form were informed by theory (Linacre, 1993). Empirical estimation of reliability for the shorter form was hampered by the systematic omission of one-third of the items from each of the two survey forms. A selection of 25 items, for instance, averaged fewer than 18 items per respondent. Accordingly, after examining 25-, 29-, and 30-item solutions, the 29-item version of the scale was selected as exhibiting the desired properties.

Most of the items with fit statistics over 1.70 were removed in the reduction of the scale to 29 items; for this version of the FCSS, the item mean square infit statistics ranged from 0.70 to 1.78, with point-measure correlations ranging from .60 to .76. Given the sample size and numbers of rating scale categories, these results indicate data consistency of the quality required for meaningful and useful interpretations of the measures.

Because the large number of items included in the FCSS led to the construction of two forms, the question as to the comparability of the measures from the two sets of items arises. Figure 7 shows a scatter plot of two sets of measures made from two groups of FCSS items, those unique to one form vs. those on both forms. The correlation of the pairs of measures was .95 (disattenuated, 1.00), indicating that the items included on just one form measure the same thing as the items included on both forms.

As with the IFS, separate calibrations of the FCSS items were also estimated for each demographic subsample. Overall, correlations were all greater than .78, with the vast majority over .90. The correlations of the by-state FCSS calibrations ranged from .78 to .94. Due to its low size (85), all seven of the Florida sample's correlations were in the bottom of the overall range of correlations (.78 to .84; disattenuated, .88 to .95). The other 21 of the 28 were all greater than .88; disattenuated, these 21 ranged from .91 to 1.00. The three correlations between separately estimated FCSS calibrations for the three different subsamples based in child ages (0-1, 1-2, and 2-3) ranged from .95 to .97. The calibrations by each of the five main ethnic groups correlated .84 to .96 (.95 to 1.00, disattenuated). The correlation of the FCSS calibrations produced by data from families who filled out the form independently with those produced by families who had the survey read to them was .99. The FCSS calibrations correlated .98 when estimated from the 416 persons completing the survey online vs. the 1,334 who used paper forms, and they correlated .85 (.92, disattenuated) for those responding to the English language form (1,415 measured cases) vs. those responding to the Spanish language form ($n = 178$).

Scale Reliability

Reliability is traditionally defined as the proportion of the total variance that can be characterized as true, independent of the effect of random error. This proportion can be expressed in terms of the separation statistic, G , which indicates the spread of items or persons in standard error units (Wright, 1996; Wright & Masters, 1982). The availability of individual error estimates for each person measure and item calibration makes it possible to subtract the mean square error from the variance for a direct estimate of the true variance. The square root of the true variance is then an adjusted standard deviation. The ratio of that adjusted standard deviation to the root mean square error is the separation reliability. The higher the variance of the person or item locations on the measurement continuum, relative to the error, the higher the reliability of the distinctions implied by those locations.

IFS. The separation reliability of .93 found for the 1,750 IFS measures is associated with a ratio of true variation to error of about 4/1. Respondents can be separated into approximately five statistically distinct levels, or strata, along

the measurement continuum (Wright & Masters, 1982). Cronbach's alpha (KR20) for the raw scores was 1.00 for the IFS measures. Separation reliability is expected to be more conservative than Cronbach's alpha because individualized error estimates in the Rasch model provide a more precise value for the explained variance than the group-level correlations used in calculating alpha.

Separation reliability for the IFS item calibrations was 1.00, indicating that the item locations can be separated into approximately 18 statistically distinct ranges along the measurement continuum.

FCSS. The separation reliability of .94-.97 found for the 1,755 FCSS measures on the full 135-item scale imply a ratio of true variation to error of at least 4/1. Respondents separate into about four strata on the shorter 29-item FCSS ($r = .87-.92$). Cronbach's alpha (KR20) for the respondents' sum scores was 1.00 for the FCSS measures. Separation reliability for the FCSS item calibrations was 1.00, indicating that the item locations can be separated into approximately 19 strata. Because the respondent sample separates the better targeted and better fitting 29-item FCSS with slightly lower error relative to increased variation, 22 strata are distinguished.

Dimensionality

As shown in the scree plot in Figure 8, a Principal Components Analysis (PCA) of the IFS scores performed using SPSS indicated that about 68% of the variance was explained by the first component dimension (using listwise deletion; the percentage dropped to 66% with pairwise deletion, and 63% with mean score replacement). The second component had an eigenvalue of 1.3 and accounted for 4.6% of the total variance.

A PCA of the IFS standardized residuals (Linacre, 1998, 2003a, 2008; Smith, 1996a; Wright, 1996) performed in Winsteps (Linacre, 2010) approximately reproduced these results, finding 60% of the total variance explained by the primary dimension, with the first residual factor accounting for 4.6% of the total variance. The two groups of seven and nine items at the opposite ends of the first contrast with loadings greater than $|.20|$ were used to produce two sets of measures; reliabilities were both about .85, and the measures correlated .82 (1.00, disattenuated). Overall, these results indicate that the scale is essentially unidimensional (Linacre, 2003a; R. Smith, 1996a; E. Smith, 2002).

Similar results were obtained for the FCSS, with the Winsteps PCA showing about 54% of the variance explained by the primary dimension, and the first residual factor accounting for 3.5% of the total variance.

Prediction of IFS Based on FCSS

The quality of family-centered EI services can plausibly be considered the primary cause of the impacts of those services on families. EI service impacts are likely also shaped to varying extents by the children and families served, the educational backgrounds and skill levels of the program coordinators and service providers, as well as by the program leadership, funding, and community support (Talan & Bloom, 2004). Though the limited scope of the present data must be acknowledged, there is practical value in establishing even provisionally the magnitude of service quality impacts on families. To investigate the relationship between quality of family-centered services and impacts on families, measures on the IFS were regressed on measures on the FCSS separately for the 135-item and the 29-item versions of the FCSS.

Using FCSS measures from the entire 135-item scale, there are 1744 cases with both measures (see Table 5). The coefficient of correlation is .83, indicating a strong and positive relationship between the quality of family-centered EI services and the impacts of those services on families. As expected, families that feel they receive higher quality EI services tend to report a greater impact from them. The line of best fit explains about 70% of the variation in the IFS measures ($R^2=0.70$, Adj. $R^2=.69$).

As shown in Table 6, after taking the reduced precision of the shorter scale into account, approximately the same results were obtained for the 29-item FCSS measures.

Standard Setting

A nationally representative group of stakeholders in early intervention was convened to recommend measures on the IFS that would represent a minimum standard for each of the Indicator C-4 sub-indicators. Standards were not set for the FCSS, as the FCSS, while providing valuable information for improvement purposes, does not directly address outcomes of early intervention services for the family.

The three major components of a standard setting task in their original context of educational testing have been described by Stone (2001). Fisher, Elbaum, and Coulter (2010) have described these three tasks as modified by NCSEAM in the context of states' federal reporting requirements.

The diagram shown in Figure 9, known as a Wright map (Wilson, 2005), was employed in the standard-setting session, and serves to illustrate how the IFS was used in the standard setting process. (See Figure 10 for a Wright map of the FCSS.) The question was raised as to how far up the scale a family's measure would have to be before it could be said that a given early intervention program was satisfying the terms of each of the outcome areas. For example,

could early intervention services be claimed to have helped the family “help their children develop and learn” if families feel that their efforts are helping their child (calibrating at 498), but they do not endorse items located further up the scale? In this case, the stakeholder group concurred that families would minimally have to report that early intervention services helped them “understand their child’s special needs” (calibrating at 516). Stakeholders also agreed that in order to fulfill the intent of the outcome related to families knowing their rights, families would have to endorse the item with that specific content (“Over the past year, early intervention services have helped me and/or my family know about my child’s and family’s rights concerning early intervention services” (calibrating at 539). Finally, for the outcome related to communication, stakeholders set the minimum standard at the location on the scale corresponding to the item, “Over the past year, early intervention services have helped me and/or my family communicate more effectively with people who work with my child and my family” (calibrating at 556).

After unanimously agreeing on the vertical levels at which horizontal lines were to be drawn—just above items 18 (for outcome area A), 24 (for outcome area B), and 15 (for outcome area C) in Figure 1—stakeholders participating in the standard-setting session were requested, in the second step, to draw a vertical line through the horizontal line at the point at which the probability of an agreeable response was sufficiently high for them to feel confident in the degree of agreement expressed. The probability chosen was a 0.95 probability of agreement.

The final unanimous decision concerned whether the standard should be interpreted as taking effect at the indicated measures, at the lower bound of the 95% confidence interval, or at least at those measures, using the upper bound. For simplicity’s sake, it was agreed that the cut-off values would be set at the item calibration values. As users’ facility with the measurement system increases, the more sophisticated methods that take the error of measurement into account might be adopted.

Applying the recommended standards, how do states derive the percent on each of the sub-indicators? Since there is a single distribution of measures for any given data collection, the percent on any of the sub-indicators is calculated as the percent of families with measures at or above the designated standard. Given that the standard for sub-indicator B is set further up the scale than the standard for, say sub-indicator C—meaning that early intervention services have to have a greater positive impact on family outcomes in order for families to know their rights than for families to help their child develop and learn—the percent reported for sub-indicator B will always be smaller than the percent reported for sub-indicator C. This is not an artifact of the instrument but a reflection of the fact that it is more challenging to achieve certain family outcomes than it is to achieve others.

Discussion

The consistency of data across multiple independent samples of respondents provides compelling evidence that the IFS and FCSS measure distinct unidimensional constructs. Beyond having appropriate item content, a measurement tool should also yield invariant measures, ones that mean the same thing for all families who provide their responses. If item response patterns are different for different subgroups of respondents—for example, for families of younger children versus families of older children, or for families of Hispanic origin versus families of African-American origin—that finding invalidates comparisons of the measures across groups. The demonstration of invariance for different groups of respondents, not in the absolute level of the trait being measured, but in the pattern of responses, provides the strongest support for the validity of the measures.

The reliabilities of the IFS and FCSS measures in this study were very high. Reliabilities for psycho-educational assessments are widely considered acceptable if they exceed .8, and excellent if they exceed .9. High-stakes educational examinations are often held to reliability standards of .95 and higher. Therefore, the IFS and FCSS meet the most stringent standards of measurement reliability that are applied in the social sciences.

The strong reliability and validity evidence for the IFS supports its use in federal accountability applications as well as the interpretation of IFS measures as representing consistent and uniform differences in the probability of agreement with statements concerning the impact of EI services on families. Tables 7 and 8 provide guidelines for the interpretation of the IFS and FCSS measures.

As of 2009, approximately 25 states have been using the IFS to collect data addressing Indicator C4 of the State Performance Plan (Office of Special Education Programs, 2010). Several state lead agencies for EI (e.g., the New York State Bureau of Early Intervention, NYSBEI) have also been using the FCSS. As part of a General Supervision Enhancement Grant awarded to the NYSBEI, that agency piloted use of a 25-item FCSS on a sample of 230 families exiting the early intervention program. For that sample, the FCSS had a .98 measurement reliability, as estimated through the Rasch model.

Results of our analyses indicated that there is a strong, positive association between measures on the FCSS and measures on the IFS. The implication of this finding is that improvements in FCSS are likely to be strongly associated with increases in IFS measures, and improved percentages on the accountability indicators. Consider, for instance, a family with an FCSS measure of about 330, almost two standard deviations below the mean in the study sample. This family disagrees with most of the FCSS items and agrees only that their service

providers are good at working with families, knowledgeable and professional, easy to talk to, and dependable. As shown in Tables 3 and 8, and Figure 10, this family's measure is interpreted as saying that the written information they receive is not sufficiently clear or understandable, their schedule is not taken adequately into account in planning meetings, and that staff from the EI program had not visited their home to give them ideas on helping their child. This family would accordingly not be expected to agree very strongly with even the most agreeable items on the IFS scale (Tables 2 and 7, and Figure 9), that EI services have helped them do things for and with their child that are good for the child's development, made them feel their efforts are helping their child, or helped them better understand their child's needs, or be more optimistic about their child's future.

Alternatively, should the program improve the family-centeredness of services provided to this family, a subsequent administration of the FCSS could show a marked improvement, perhaps to a measure closer to the mean, such as 475. At this level, the family would be agreeing strongly or very strongly with items at the bottom of the scale, with which they had previously expressed only reserved agreement. And they would now also be agreeing with items further up the scale, such as items indicating that they have felt like part of the EI team, the information they receive is understandable, and that the family's needs were considered when planning for the child's services. Now this family would reasonably be expected to also agree with items further up the IFS scale, indicating, for example, that the family had experienced positive impacts such as feeling more competent to handle the challenges of parenting a child with special needs, being able to communicate more effectively with people who work with their child and the family, and understanding how the early intervention system works. Linking the quality of service delivery to family outcomes represents a powerful tool for overall program improvement and, consequently, for achieving improved outcomes for young children with disabilities and their families.

Conclusion

The results of these analyses support the conclusion that measures on the IFS can be interpreted as quantifying the impact of early intervention programs on families. Statewide distributions of families' measures on the IFS, based on a representative sample of families participating in early intervention, provide a sound foundation for reporting the percent of measures that are above the standard for each of the outcome areas. These percentages can be interpreted as a valid and reliable index of "the percent of families who report that early intervention services have helped the family (a) know their rights; (b) effectively

communicate their children's needs; and (c) help their children develop and learn."

The content and hierarchical order of the FCSS items delineate a methodical path toward improved program impacts on families. With about 70% of the variance in the IFS measures accounted for by differences in the FCSS measures, measurable improvements in program quality statewide are likely to be associated with increases in the annual IFS measures used in calculating the Part C SPP/APR Indicator 4 percentages. Increases in these percentages in turn indicate improved outcomes for families.

References

- Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-574.
- Andrich, D. (1996). Measurement criteria for choosing among models with graded responses. In A. von Eye & C. Clogg (Eds.), *Categorical variables in developmental research: Methods of analysis* (pp. 3-35). New York: Academic Press, Inc.
- Andrich, D. (2009). Understanding the response structure and process in the polytomous Rasch model. In M. Nering & R. Ostini (Eds.), *Handbook of polytomous Item Response Theory models: Developments and applications* (pp. 123-152). Lawrence Erlbaum Associates, Inc.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*, 2nd ed. Mahwah, NJ: Erlbaum.
- Data Accountability Center (2011). Table 6-7: Infants and toddlers receiving early intervention services under IDEA, Part C, by race/ethnicity and state: Fall 2005. Retrieved from https://www.ideadata.org/arc_toc7.asp#partcCC
- Fisher, W. P., Jr., Elbaum, B., & Coulter, W. A. (2010). Stakeholder input in scale development and standard setting: An example from the federal accountability system for special education. Manuscript submitted for publication.
- Fisher, W. P., Jr., & Wright, B. D. (Eds.). (1994). Applications of probabilistic conjoint measurement. *International Journal of Educational Research*, 21(6), 557-664.
- IDEA. (2004). Individuals with Disabilities Education Improvement Act, Amendments of 2004, Pub. L. No. 108-446, U.S.C. 20, §1400 et seq.
- Linacre, J. M. (1993). Rasch-based generalizability theory. *Rasch Measurement Transactions*, 7(1), 283-284.
- Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, 2(3), 266-83.
- Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, 3, 103-122.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre, J. M. (2003a). Data variance: Explained, modeled and empirical. *Rasch Measurement Transactions*, 17(3), 942-3.
- Linacre, J. M. (2003b). Size vs. significance: Infit and outfit mean-square and standardized chi-square fit statistics. *Rasch Measurement Transactions*, 17(1), 918.
- Linacre, J. M. (2004). Optimizing rating scale category effectiveness. In E. V. Smith, Jr. and R. M. Smith (Eds.), *Introduction to Rasch measurement:*

- Theory, models, and applications* (pp. 258-278). Maple Grove, MN: JAM Press.
- Linacre, J. M. (2008). Variance in data explained by Rasch measures. *Rasch Measurement Transactions*, 22(1), 1164.
- Linacre, J. M. (2010). WINSTEPS Rasch measurement software [Computer program], version 3.70. Chicago: WINSTEPS.com.
- Luo, G. (2005). The relationship between the rating scale and partial credit models and the implication of disordered thresholds of the Rasch models for polytomous responses. *Journal of Applied Measurement*, 6(4), 443-55.
- Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, 56(1), 63-75.
- Regional Resource Center Program. (2008). Indicator Support Grid. Retrieved from <http://spp-apr-calendar.rrfcnetwork.org/getfile/view/id/139>.
- Schumacker, R. E. (1996). Disattenuating correlation coefficients. *Rasch Measurement Transactions*, 10(1), 479.
- Smith, E. V., Jr. (2002). Understanding Rasch measurement: Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3(2), 205-31.
- Smith, R. M. (1996a). A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling*, 3(1), 25-40.
- Smith, R. M. (1996b). Polytomous mean-square fit statistics. *Rasch Measurement Transactions*, 10(3), 516-517.
- SPSS. (2003). SPSS for Windows, Rel. 12.0.0. Chicago: SPSS Inc.
- Stenner, A. J. (1995). Point-biserial fit indices. *Rasch Measurement Transactions*, 9(1), 416.
- Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2006). How accurate are Lexile text measures? *Journal of Applied Measurement*, 7(3), 307-22.
- Stone, G. (2001). Understanding Rasch measurement: Objective standard setting (or truth in advertising). *Journal of Applied Measurement*, 2(2), 187-201.
- Talan, T., & Bloom, P. J. (2004). Program administration scale: Measuring early childhood leadership and management. New York: Teachers College Press.
- U. S. Office of Special Education Programs (2010). Part C State Performance Plan/Annual Performance Report 2010 indicator analyses. Retrieved from http://leadershipmega-conf-reg.tadnet.org/uploads/file_assets/attachments/204/original_original_part-c_sppapr_08-rev_1_.pdf?1280109704
- Wright, B. D. (1991). Errors, variances and correlations. *Rasch Measurement Transactions*, 5(2), 147.

- Wright, B. D. (1996). Comparing Rasch measurement and factor analysis. *Structural Equation Modeling*, 3(1), 3-24.
- Wright, B. D., & Masters, G. N. (1982). Rating scale analysis: Rasch measurement. Chicago, Illinois: MESA Press.
- Wright, B. D., & Masters, J. N. (1982). Rating scale analysis. Chicago: MESA Press.
- Wright, B. D., & Stone, M. H. (1999). Measurement essentials. Wilmington, DE: Wide Range, Inc.

Table 1

Sample Demographics

Characteristic	<i>N</i>	Percent of Sample	Percent of U.S. Population Ages 0-3 Receiving Early Intervention Services
State of residence			
California	262	14.4	
Florida	85	4.7	
Georgia	144	7.9	
Iowa	287	15.8	
Louisiana	143	7.9	
Massachusetts	422	23.2	
New Jersey	164	9.0	
New Mexico	301	16.6	
Unknown	8	.4	
Ethnicity			
White	1,041	57.3	60.4
Black or African American	150	8.3	13.8
Hispanic or Latino	394	21.7	20.4
Asian or Pacific Islander	70	3.9	4.4
American Indian or Alaskan Native	93	5.1	1.0
Multi-Racial	51	2.8	n/a
Unknown	17	0.9	
Areas in which child has special needs (multiple responses allowed)			
Understanding and using language	1,277	70.3	

Learning and cognition	745	41.0
Social skills / Behavior	578	31.8
Emotional	274	15.1
Adaptive skills	409	22.5
Physical / Movement	827	45.5
Health / Medical	548	30.1
Child's age at time of survey		
0-6 months	51	2.8
6-12 months	171	9.4
12-18 months	211	11.6
18-24 months	286	15.7
24-30 months	421	23.2
30-36 months	447	24.6
Over 3 years old	222	12.2
Unknown	7	.4
Child's age upon referral to EI		
0-6 months	632	34.8
6-12 months	322	17.7
12-18 months	311	17.1
18-24 months	264	14.5
24-30 months	182	10.0
30-36 months	44	2.4
Over 3 years old	11	0.6
Unknown	50	2.8
Form administered		

Form 1 (101 items)	982	54.1
Form 2 (100 items)	834	45.9
Mode of administration		
Completed the survey independently	1,435	79.0
Completed the survey while facilitator read items aloud	267	14.7
Unknown	114	6.3
Language in which the survey was completed		
English	1,474	81.2
Spanish	178	9.8
Another language	11	.6
Unknown	153	8.4

Note: Data on state of residence, ethnicity, areas in which the child has special needs, child's age at time of survey, child's age upon referral to EI, and mode of survey administration are based on parents' responses on the survey forms.

Table 2

Full Text of IFS Items

Item	Item text
calibration	Over the past year, early intervention services have helped me and/or my family:
678	participate in typical activities for children and families in the community.
656	know about services in the community.
640	know where to go for help and support to meet my family's needs.
625	keep up friendships for my child and family.
609	know where to go for help and support to meet my child's needs.
606	cope with stressful situations.
600	help other children in my family adjust to their brother's or sister's special needs.
584	be more effective in managing my child's behavior.
580	find information I need.
576	make changes in our family routines that will benefit my child with special needs.
576	do activities that are good for my child even in times of stress.
570	improve my family's quality of life.
565	feel that I can get the services and supports that my child and family need.
563	get services that my child and family need.
562	feel that my family will be accepted and welcomed in the community.
559	feel more confident in my skills as a parent.

- 559 feel that my child will be accepted and welcomed in the community.
- 556 communicate more effectively with people who work with my child and my family.
- 553 understand how the early intervention system works.
- 546 understand the roles and responsibilities of the people who work with my child and family.
- 540 figure out solutions to problems as they come up.
- 539 feel that I can handle the challenges of parenting a child with special needs.
- 539 know about my child's and family's rights concerning early intervention services.
- 534 be able to evaluate how much progress my child is making.
- 530 be more optimistic about my child's future.
- 516 understand my child's special needs.
- 498 feel that my efforts are helping my child.
- 498 do things with and for my child that are good for my child's development.
-

Table 3

Full Text of FCSS Items (29-item version)

Item calibration	Item text
631	Someone from the EIP went out into the community with me and my child (to a park, church, child care facility, etc.) to help me get involved in community activities and services.
600	The early intervention program regularly holds public meetings to to gather family input on early intervention services.
581	My family was asked whether I wanted help in dealing with stressful situations.
570	Someone from the EIP helped me get in touch with other parents for help and support.
559	My family was given information about community programs that are open to all children.
556	My family was given information about organizations that offer information and training for parents, for example, Parent Training and Information Centers, Family Resource Centers, disability support groups, etc.
553	Someone from the EIP helped me get services like child care, transportation, respite care, pre-school programs, WIC/Food stamps, etc.
550	My family was given information about how to access different programs and services in the community.
524	My family was given information about opportunities for my child to play with other children.
513	My family was given information about how to advocate for my child and my family.
512	My family was given information about who to call if I am not satisfied with the services my child receives.

- 510 I was given information to help me prepare for my child's transition.
- 503 I was offered help I needed, such as child care services or transportation, to enable me to participate in the IFSP meeting(s).
- 494 My child receives the amount of therapy sessions that he/she needs.
- 483 I was given choices concerning my family's services and supports.
- 482 The services on our IFSP have been provided in a timely manner.
- 477 My family was given information about my rights as a parent of a child who is eligible for early intervention services.
- 477 My family was given information about modifications of routines, activities, and the physical setting that would help my child in different environments.
- 472 My family's needs (such as transportation, child care, etc.) were considered when planning for my child's services.
- 470 Someone from the EIP visited my home to give me ideas on helping my child at home.
- 470 The written information I receive is clear to me.
- 462 My service coordinator is available to speak with me on a regular basis.
- 454 The written information I receive is in a language I understand.
- 452 My family's schedule and daily routines were considered when planning for my child's services.
- 444 I have felt part of the team when meeting to discuss my child.
- 430 The early intervention service providers that work with my child are dependable.
- 429 My service coordinator is good at working with families.
- 424 The early intervention service providers that work with my child are easy for me to talk to about my child and family.

418 The early intervention service providers that work with my child are knowledgeable and professional.

Table 4

Summary Statistics for the IFS and FCSS (135-item Version and 29-item Version)

	IFS (28 items)	FCSS (135 items)	FCSS (29 items)
Sample Sizes	N-E measures: 1,572	N-E measures: 1,556	N-E measures: 1,452
	All measures: 1,750	All measures: 1,755	All measures: 1,755
Measure	N-E: 615 / 112 / 23	N-E: 534 / 102 / 14	N-E: 504 / 91 / 25
Mean/SD/Error	All: 644 / 158 / 31	All: 562 / 139 / 23	All: 541 / 129 / 37
Mean Item Count & Respondent Score	N-E: 26.3 / 95.5 All: 26.2 / 98.0	N-E: 91.3 / 220.2 All: 91.4 / 225.3	N-E: 20.4 / 46.5 All: 20.3 / 48.6
Measurement Separation G & Reliability (Cronbach/ Rasch)	N-E Real: 4.5 / 1.00 / .95 N-E Modeled: 5.1 / .96 All Real: 3.7 / .93 All Modeled: 3.8 / .93	N-E Real: 5.5 / .99 / .97 N-E Modeled: 6.2 / .97 All Real: 3.8 / .94 All Modeled: 3.9 / .94	N-E Real: 2.9 / .95 / .89 N-E Modeled: 3.3 / .92 All Real: 2.6 / .87 All Modeled: 2.7 / .88
Calibration Mean/SD/Error	570 / 43 / 3	500 / 50 / 3	500 / 55 / 3
Mean Count of Respondents & Item Sum Score	1639.7 / 6125.6	1187.6 / 2928.8	1231.5 / 2939.7
Calibration Mean Square Fit Mean/SD	Infit: 1.00 / .18 Outfit: 1.03 / .23	Infit: 1.02 / .34 Outfit: 1.09 / .69	Infit: 1.05 / .29 Outfit: 1.14 / .42
Calibration Separation G & Reliability	Real: 11.0 / .99 Modeled: 11.4 / .99	Real: 13.2 / .99 Modeled: 14.2 / 1.00	Real: 15.1 / 1.00 Modeled: 16.1 / 1.00

Note: N-E = non-extreme. Cronbach's alpha is shown only once as it is estimated only once, on the basis of the respondent scores.

Table 5

Regression of IFS Measures on 135-item FCSS measures (n=1,744)

Variable	B	SE B	β	<i>t</i>	Sig.
Constant	364.799	4.898		74.481	.000
FCSS	4.310	.068	.833	62.941	.000

Note. $R^2 = .70$; Adjusted $R^2 = .69$.

Table 6

Regression of IFS Measures on 29-item FCSS Measures (n=1,744)

Variable	B	SE B	β	<i>t</i>	Sig.
Constant	362.775	5.063		71.658	.000
FCSS	4.643	.076	.826	61.157	.000

Note. $R^2 = .68$; Adjusted $R^2 = .68$.

Table 7

IFS Interpretation Guidelines

Stratum	Measurement Range	Mean Error	Expected Responses	Amount of Impact on Families Measured
6	780 to 900	25-110	Very Strong Agreement with all survey items.	Full impact in all areas
5	680 to 780	16-23	Agree with items over 615; Strongly Agree with items over 500; Very Strongly Agree with items below 500.	Superior impact on fundamentals, significant impact on stress and participation in community
4	590 to 680	17-16	Agree with items over 550; Strong Agreement with items below 550.	Significant impact on fundamentals to communication with providers, feeling accepted in community
3	500 to 590	19-17	Agree with items over 515; Strongly Agree with items below 515.	Moderate impact beyond fundamentals to accessing services, dealing with stress, community activities
2	420 to 500	17-19	Disagreement on items over 570; Agreement with items below 570.	Small impact in most elementary areas of feeling effective in efforts, understanding, and optimism.
1	100 to 420	110-16	Disagreement with all or almost all survey items.	No measurable impact.

Table 8

FCSS Interpretation Guidelines

Stratum	Measurement Range	Mean Error	Expected Responses	Amount of Impact on Families Measured
4	630 to 780	29-93	Very Strong or Strong Agreement with all items.	Highest measurable service quality and family partnering efforts.
3	520 to 630	18-26	Agree with items over 530; Strong Agreement with items up to 530.	Significant to superior service quality and family partnering efforts.
2	440 to 520	16-17	Disagree with items over 570; Agree with items above 530; Strong Agreement with items calibrating at 530 or lower.	Moderate service quality and family partnering efforts.
1	135 to 440	92-16	Disagree with items calibrating at 530 or higher; agreement with those below 530.	Low or no measurable service quality or family partnering efforts.

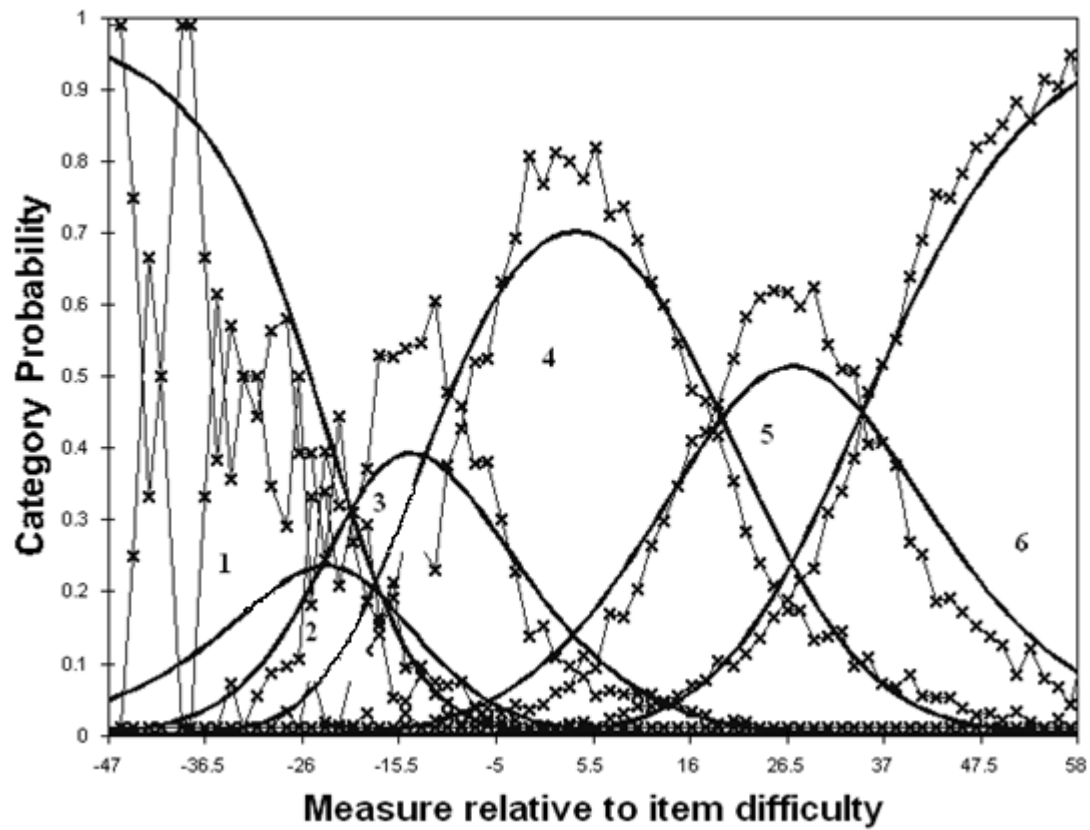


Figure 1. Response Probabilities for Impact on Families Scale, Original Six Categories

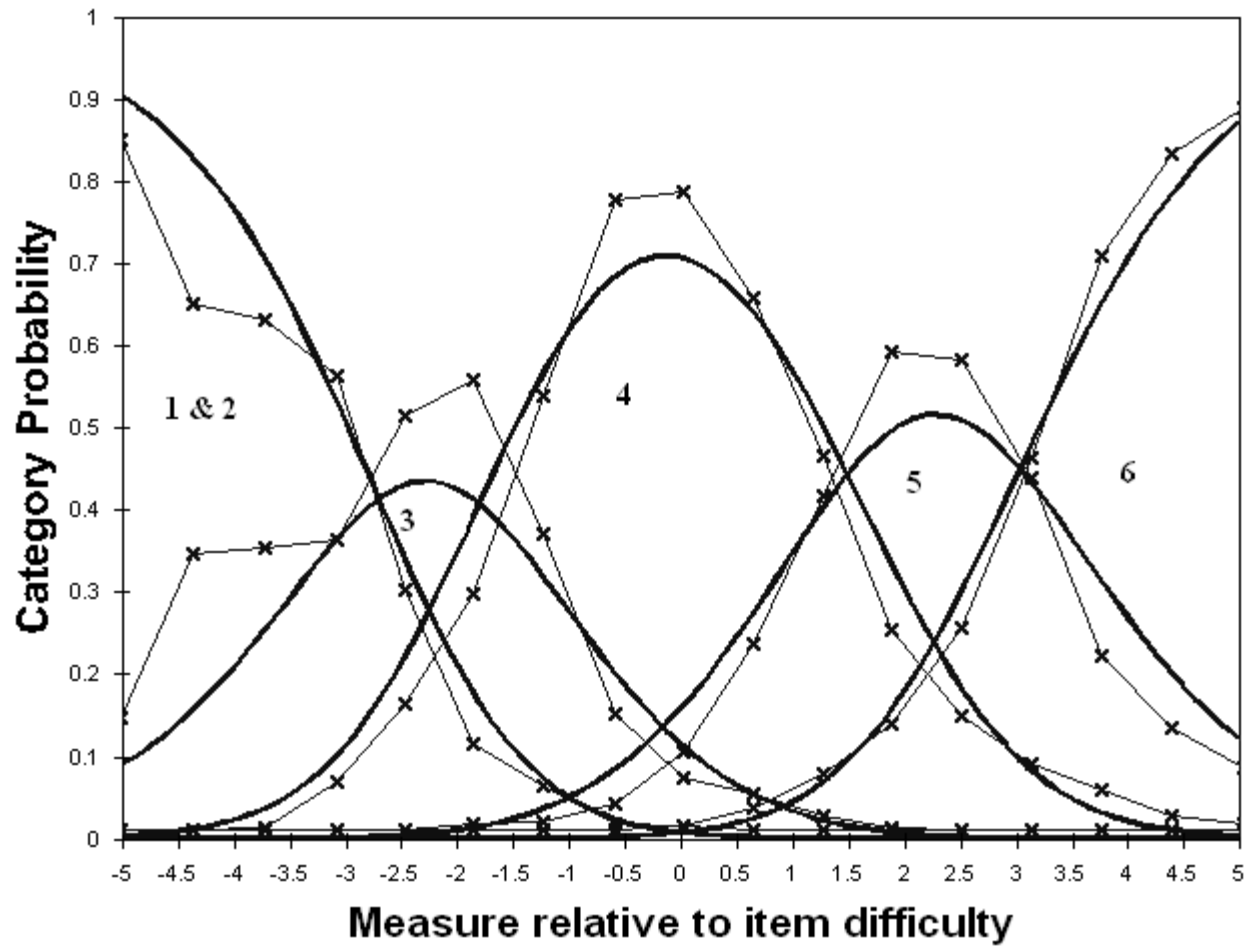


Figure 2. Response Probabilities for Impact on Families Scale, Optimized Five Categories

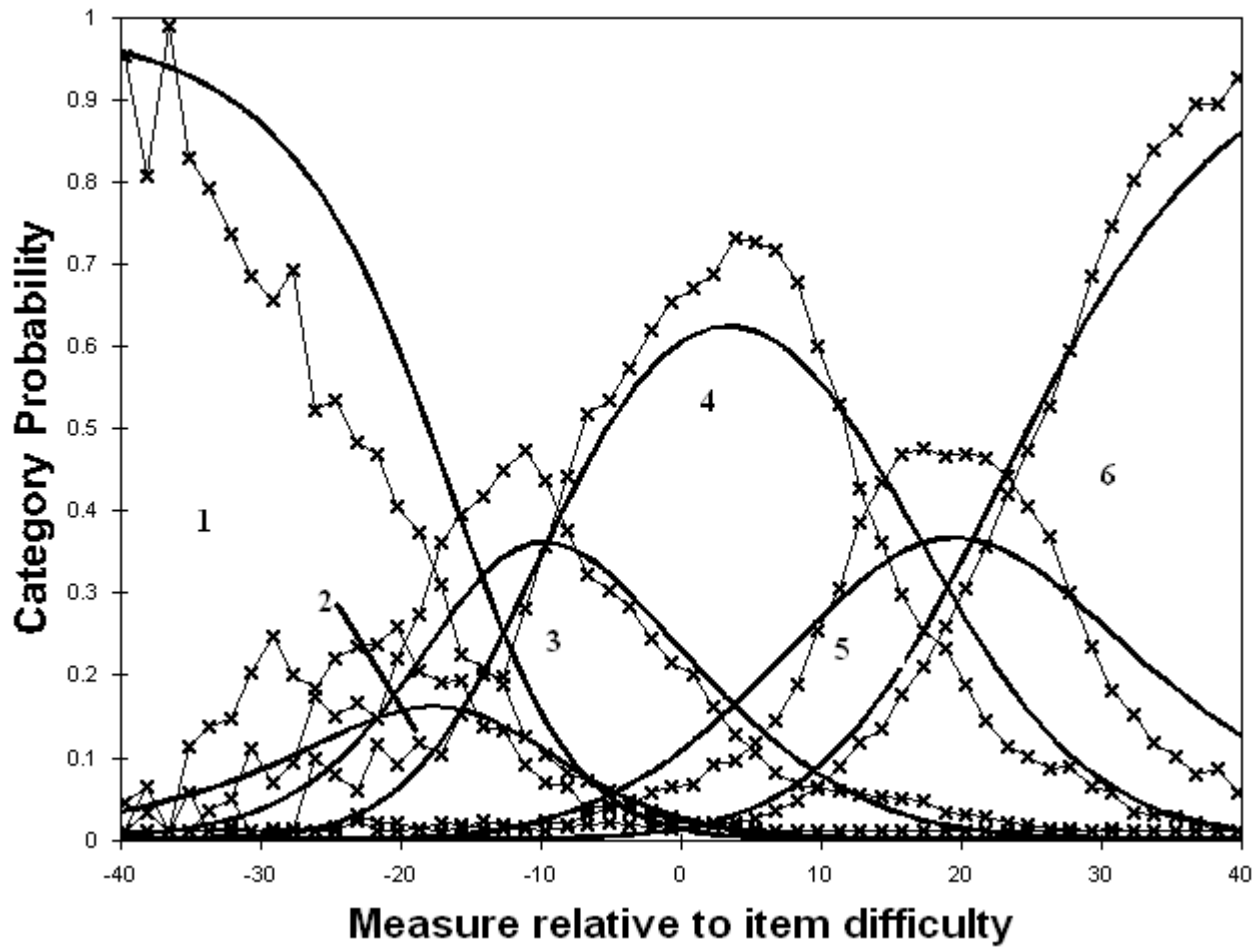


Figure 3. Response Probabilities for Family Centered EI Services Scale, Original Six Categories

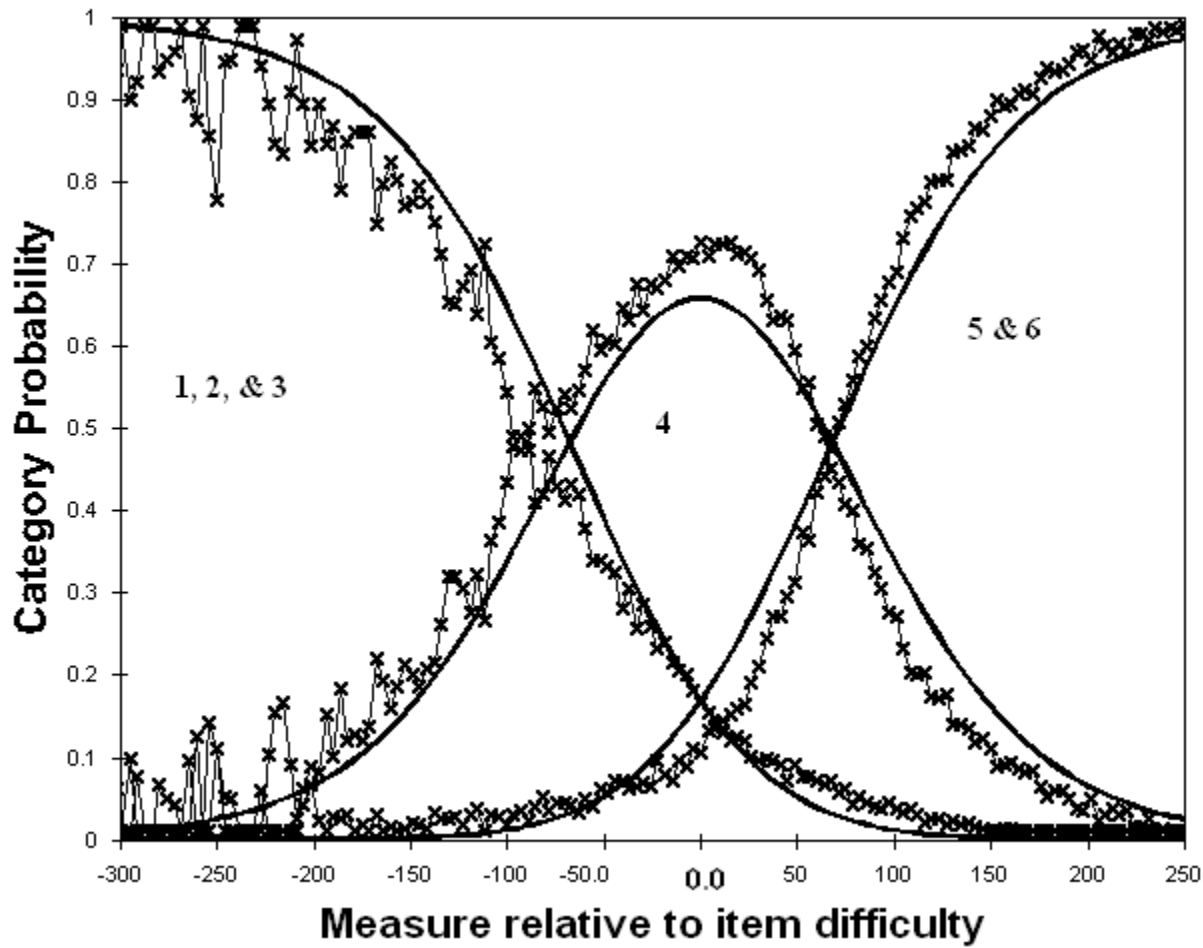


Figure 4. Response Probabilities for Family Centered EI Services Scale, Optimized Three Categories

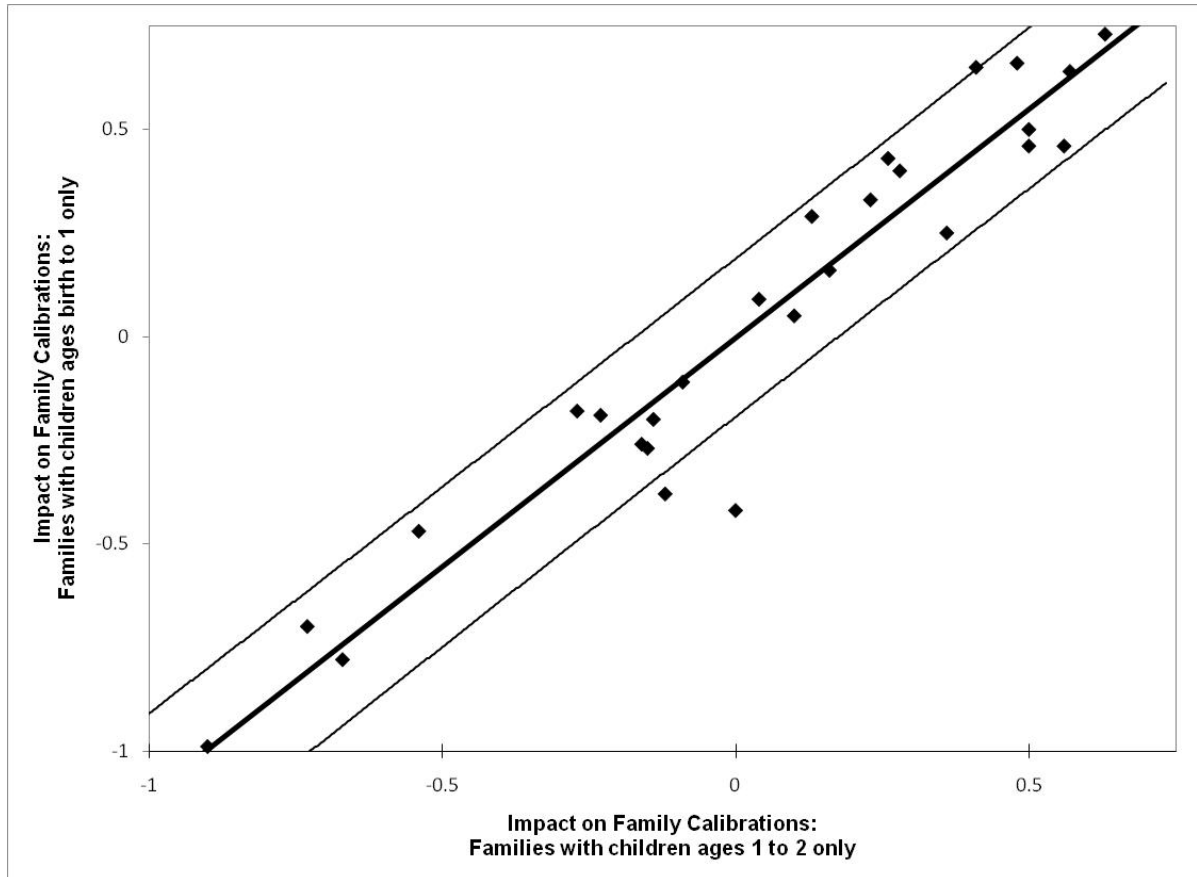


Figure 5. Scatter plot of IFS item calibrations: families of children age 0-1 vs. families of children ages 1-2 ($r = .96$; disattenuated, 1.00)

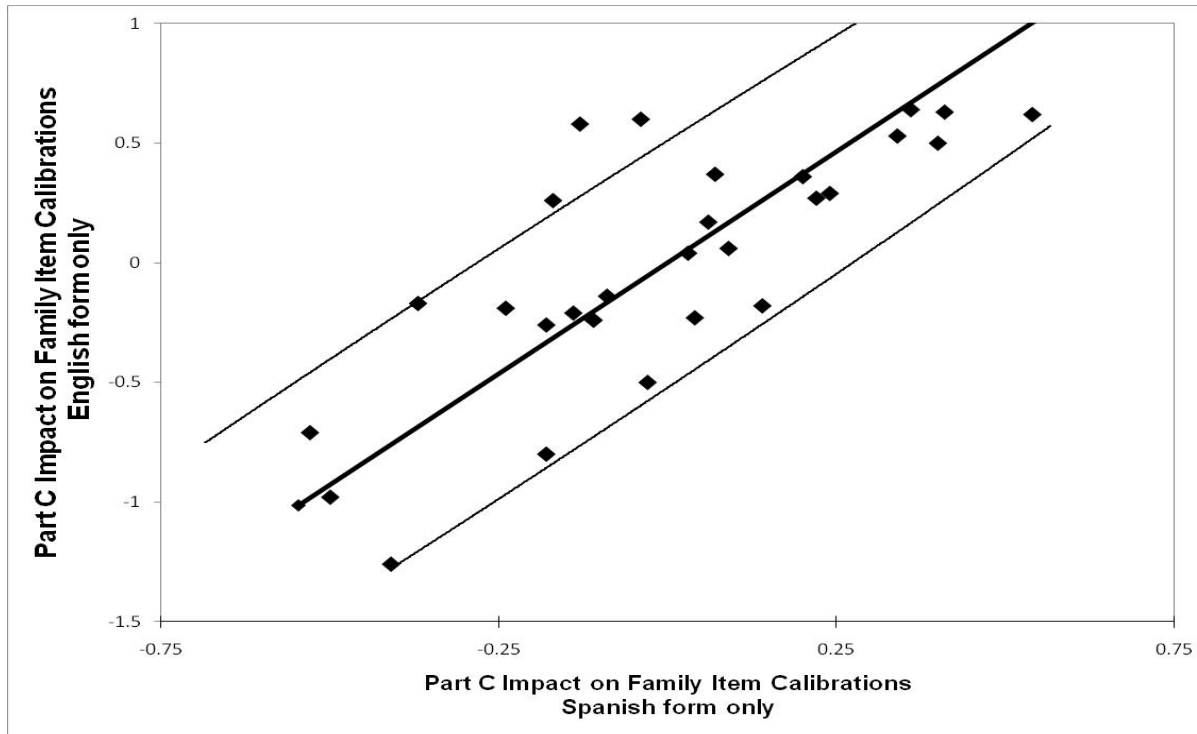


Figure 6. Scatter plot of IFS item calibrations: English-language vs. Spanish-language ($r = .78$; disattenuated, 1.00)

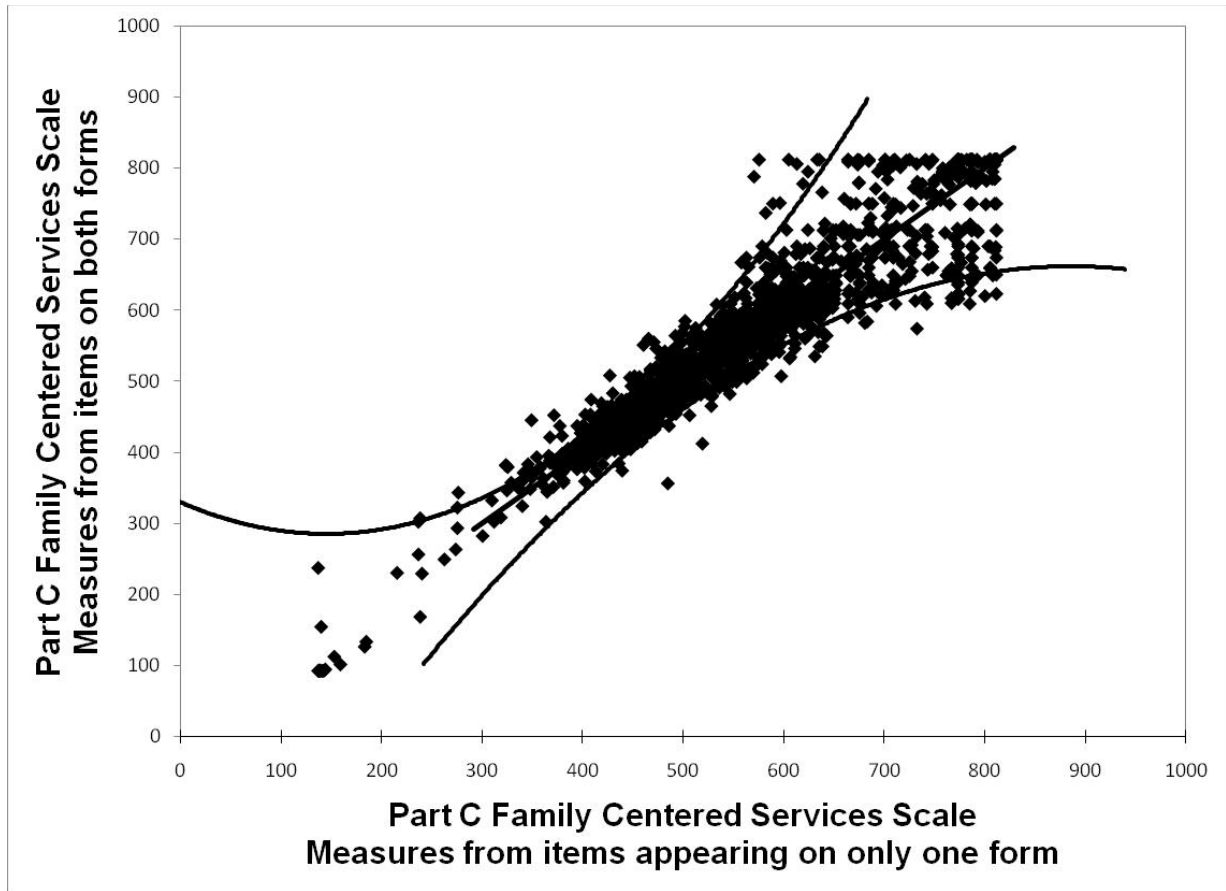


Figure 7. Scatter plot of FCSS measures: items unique to one form vs. those on both forms; $r = .95$; disattenuated, 1.00)

Scree Plot

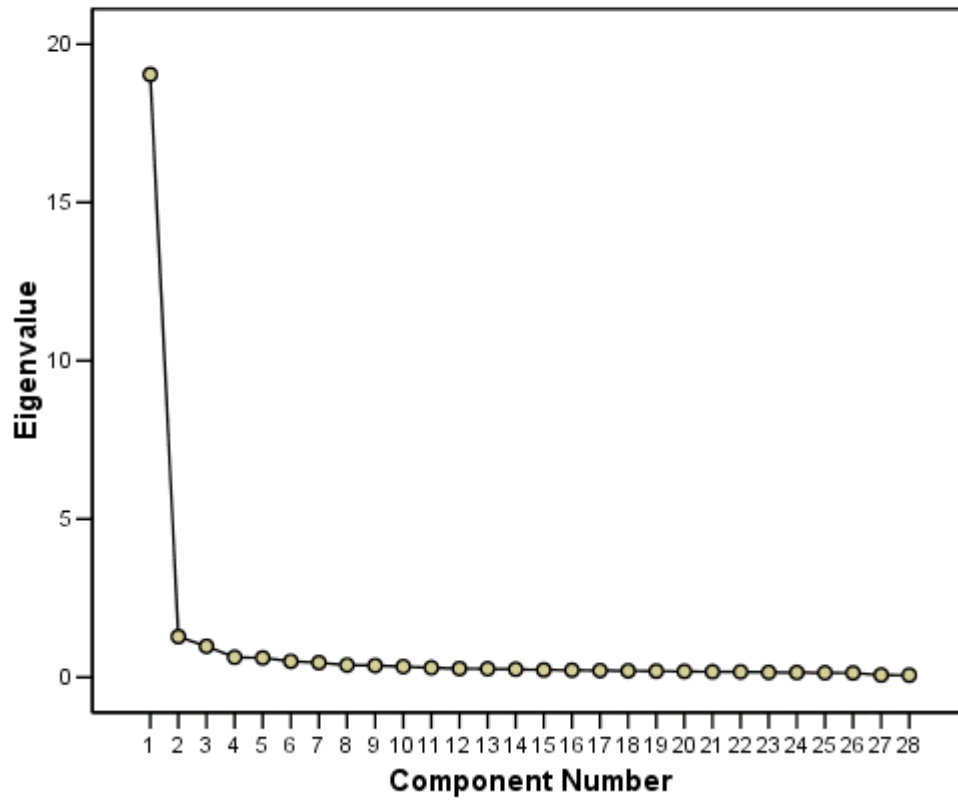


Figure 8. Scree Plot for Unrotated Principal Components Analysis of the Impact on Family Scale (IFS)

```

EXPECTED SCORE: MEAN (Rasch-score-point threshold, ":" indicates Rasch-half-point threshold) (BY CATEGORY SCORE)
100 200 300 400 500 600 700 800 900 1000
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----|
1 LEAST 1 : 2 : 3 : 4 : 5 5 NUM ITEM Over the past year, early intervention services have helped me and/or my family:
1 AGREEABLE 1 : 2 : 3 : 4 : 5 5 35 participate in typical activities for children and families in the community.
1 1 : 2 : 3 : 4 : 5 5 20 know about services in the community.
1 1 : 2 : 3 : 4 : 5 5 22 know where to go for help and support to meet my family's needs.
1 1 : 2 : 3 : 4 : 5 5 31 keep up friendships for my child and family.
1 1 : 2 : 3 : 4 : 5 5 21 know where to go for help and support to meet my child's needs.
1 1 : 2 : 3 : 4 : 5 5 33 cope with stressful situations.
1 1 : 2 : 3 : 4 : 5 5 26 help other children in my family adjust to their brother's or sister's spec
1 VS & S 1 : 2 DIS- 3 4 S 5 VS 5 28 be more effective in managing my child's behavior.
1 DISAGREE>1 >> 2 AGREE 3 AGREE 4 AGREE5 AGREE 5 19 find information I need.
1 1 : 2 : 3 : 4 : 5 5 27 make changes in our family routines that will benefit my child with special needs.
1 1 : 2 : 3 : 4 : 5 5 30 do activities that are good for my child even in times of stress.
1 1 : 2 : 3 : 4 : 5 5 34 improve my family's quality of life.
1 1 : 2 : 3 : 4 : 5 5 10 feel that I can get the services and supports that my child and family need.
1 1 : 2 : 3 : 4 : 5 5 32 get services that my child and family need.
1 1 : 2 : 3 : 4 : 5 5 13 feel that my family will be accepted and welcomed in the community.
1 1 : 2 : 3 : 4 : 5 5 8 feel more confident in my skills as a parent.
1 1 : 2 : 3 : 4 : 5 5 12 feel that my child will be accepted and welcomed in the community.
1 1 : 2 : 3 : 4 : 5 5 24 communicate more effectively with people who work with my child and my family.
1 1 : 2 : 3 : 4 : 5 5 16 understand how the early intervention system works.
1 1 : 2 : 3 : 4 : 5 5 17 understand the roles and responsibilities of the people who work with my child.
1 1 : 2 : 3 : 4 : 5 5 29 figure out solutions to problems as the come up.
1 1 : 2 : 3 : 4 : 5 5 11 feel that I can handle the challenges of parenting a child with special needs.
1 1 : 2 : 3 : 4 : 5 5 18 know about my child's and family's rights concerning early intervention services.
1 1 : 2 : 3 : 4 : 5 5 23 be able to evaluate how much progress my child is making.
1 1 : 2 : 3 : 4 : 5 5 9 be more optimistic about my child's future.
1 1 : 2 : 3 : 4 : 5 MOST 5 15 understand my child's special needs.
1 1 : 2 : 3 : 4 : 5 AGREEABLE 5 14 feel that my efforts are helping my child.
1 1 : 2 : 3 : 4 : 5 5 25 do things with and for my child that are good for my child's development.
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----|
100 200 300 400 500 600 700 800 900 1000
NUM ITEM

1 1\
1 1 2123462769769991755532322 2 2 4 > COUNTS OF FAMILIES MEASURED
5 1 1 2 1 141351949452103969558303532763377901212651/
T S M S T MEAN (M) AND STANDARD DEVIATIONS
0 10 20 30 50 70 80 90 99 PERCENTILE

2 2 3 3 3 4 4 6 7 8 8 9 0 2 2 3 3 3 4 \ SUMS OF
8 9 0 1 4 0 9 0 1 0 8 8 9 0 9 4 7 9 0 > RATINGS FROM
COMPLETE DATA

1 1\
0 6 4 3 2 2 1 1 1 1 1 1 1 1 2 2 3 6 0 > MEASUREMENT
8 0 3 5 6 0 7 6 7 9 8 7 6 7 1 6 6 0 8 / ERROR

```

Figure 9. Impact on Family (IFS) Wright Map 1 = VS & S Disagree, 2 = Disagree, 3 = Agree, 4 = S Agree, 5 = VS Agree

