Journal of Early Intervention
http://jei.sagepub.com/

Development and Psychometric Validation of the Family Outcomes Survey-Revised

Donald B. Bailey, Jr, Melissa Raspa, Murrey G. Olmsted, Scott P. Novak, Ann M. Sam, Betsy P. Humphreys, Robin Nelson, Nyle Robinson and Chelsea Guillen

Journal of Early Intervention 2011 33: 6
DOI: 10.1177/1053815111399441

The online version of this article can be found at:
http://jei.sagepub.com/content/33/1/6

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
DEC
Division for Early Childhood of the Council for Exceptional Children

Journal of Early Intervention
Volume 33 Number 1
March 2011 6-23

10.1177/1053815111399441
http://jei.sagepub.com
hosted at
http://online.sagepub.com

Development and Psychometric Validation of the Family Outcomes Survey–Revised

Donald B. Bailey Jr.
Melissa Raspa
Murrey G. Olmsted
Scott P. Novak
RTI International
Ann M. Sam
Betsy P. Humphreys
University of North Carolina at Chapel Hill
Robin Nelson
Texas Department of Assistive and Rehabilitative Services
Nyle Robinson
Chelsea Guillen
Illinois Department of Human Services, Division of Community Health and Prevention

Few psychometrically valid scales exist to assess family outcomes and the helpfulness of early intervention. This article describes the development and psychometric properties of the Family Outcomes Survey–Revised. The revision was prompted by the need to (a) create a new format that would be easier for parents to understand, (b) revise and expand the survey items to provide more information for states to use in planning for program improvement, and (c) demonstrate acceptable psychometric properties. Input from stakeholders and experts was used to identify concepts and develop candidate items. Data from a web-based survey conducted with 265 families in Illinois and Texas were used to assess the psychometric properties of candidate items. These activities produced a revised survey with sound psychometric integrity that can be used to document family outcomes and identify areas for program improvement.

Keywords: *family assessment; family outcomes; accountability; early intervention*

Authors' Note: Donald B. Bailey Jr., Melissa Raspa, Murrey G. Olmsted, and Scott P. Novak, Social, Statistical, and Environmental Sciences, RTI International; Ann M. Sam and Betsy P. Humphreys, School of Education, University of North Carolina at Chapel Hill; Robin Nelson, Texas Department of Assistive and Rehabilitative Services; Nyle Robinson and Chelsea Guillen, Illinois Department of Human Services, Division of Community Health and Prevention. Betsy P. Humphreys is now at the Institute on Disability, University of New Hampshire. Preparation of this article was supported in part under a cooperative agreement (H326L08001) from the Office of Special Education Programs and with support from the Texas Department of Assistive and Rehabilitative Services and the Illinois Department of Human Services. The content, however, does not necessarily represent the policy of these agencies, and you should not assume endorsement by the federal or state governments. Please address correspondence to Don Bailey, 3040 Cornwallis Road, Research Triangle Park, NC 27709-2194; e-mail: dbailey@rti.org.

Benefit to parents and families has long been considered a central goal of early intervention (Bailey et al., 1986; Dunst, 1985). However, the importance of measuring family outcomes has taken on new meaning in recent years as a result of the Government Performance Results Act of 1993. States are now required to submit annual reports to the Office of Special Education Programs (OSEP), of the U.S. Department of Education, indicating the extent to which child and family outcomes have been achieved (Bailey, Hebbeler, Olmsted, Raspa, & Bruder, 2008; Hebbeler & Barton, 2007). This requirement marks a significant departure from earlier evaluation research and reporting requirements, which focused on satisfaction with services (Bailey, Hebbeler, Spiker, Scarborough, & Mallik, 2004; Iversen, Shimmel, Ciacera, & Prabhakar, 2003; McWilliam et al., 1995) and the extent to which programs used required or recommended practices (Bailey, 2001; Bailey, Buysse, Edmondson, & Smith, 1992; Turnbull et al., 2007).

Given that the reporting of outcomes reflects a fundamental change in expectations for state-level accountability, OSEP funded the Early Childhood Outcomes Center (ECO Center) in 2003 to support national efforts to develop a set of child and family outcomes for evaluating the effectiveness of programs for infants, toddlers, and preschoolers with disabilities. At that time, there was little agreement on the most important benefits for families and the outcomes for which states should be held accountable. To address this problem, we used an evidence-based process with national stakeholder input to reach consensus on five family outcomes (Bailey, Bruder, et al., 2006). Families participating in programs for young children with disabilities should (a) understand their children's strengths, abilities, and special needs; (b) know their rights and advocate effectively for their children; (c) help their children develop and learn; (d) have support systems; and (e) access desired services, programs, and activities in their communities.

These outcomes were submitted to OSEP by the ECO Center and recommended for inclusion in future state reporting requirements. However, OSEP decided not to accept these recommendations, opting instead to ask states to document the extent to which early intervention has helped families (a) know and understand their rights, (b) communicate their children's needs, and (c) help their children develop and learn. In the remainder of this article, we refer to these three reporting requirements as *helpfulness indicators*, reflecting the degree to which parents consider early intervention to be helpful and supportive. In contrast, we use the term *outcomes* to refer to the five recommended outcomes that reflect parents' perceptions of their competence, confidence, and ability to care for their children with special needs and to achieve a satisfactory level of family adaptation.

Despite the decision to limit state reporting requirements to parent perceptions of program helpfulness, OSEP strongly encouraged ECO Center investigators to continue their work on family outcomes and develop ways for states to assess them. This encouragement was based on the assumption that federal requirements represent only the minimum level of documentation that would be expected. A more detailed measure of family outcomes would allow states to develop a more comprehensive evaluation system that uses information on program practices, family and child characteristics, family perceptions of helpfulness, and outcomes to identify areas of needed program improvement (Bailey, 2001). To help with this effort, the Family Outcomes Survey (FOS; Bailey, Hebbeler, & Bruder, 2006), a short 15-item parent-report instrument, was constructed for states to collect information on family functioning in relation to each of the five outcomes (a copy of the original

FOS can be accessed at http://www.fpg.unc.edu/~eco/assets/pdfs/FOS-PartC_11-16-06.pdf). The FOS was designed to assess family outcomes in a way that would be relatively easy for large-scale program evaluation (e.g., statewide). Scale content was developed by drawing on relevant literature, expert consultation, and an iterative process involving stakeholder feedback and revisions to the scale (Bailey et al., 2008). To assist states in meeting the OSEP reporting requirements, three questions assessed parent perceptions of whether early intervention had helped them know and understand their rights, communicate their children's needs, and help their children develop and learn.

The FOS was posted on the ECO Center website (http://www.the-eco-center.org) with open access for states, local programs, and researchers. States that had translated the survey into other languages were encouraged to submit those translations; as of this writing, 11 translations (mostly from Minnesota) have been posted on the web (Arabic, Cambodian, Croatian, Hmong, Laotian, simplified Chinese, Oromo, Russian, Somali, Spanish, and Vietnamese). According to the latest OSEP reports, at least 21 states are using the FOS, either in part or in its entirety, to assess family outcomes and report data to OSEP (Kahn, 2009).

Raspa et al. (2010) used the scale with more than 2,800 families participating in early intervention programs in Illinois and Texas. Families generally reported a high level of outcome attainment, with some variability noted across the survey items. An exploratory factor analysis revealed that outcomes clustered in two areas: (a) family knowledge and ability (the items associated with Outcomes 1–3) and (b) family support and community services (the items associated with Outcomes 4 and 5). Analyses revealed that race/ethnicity, income, time in early intervention, the perceived helpfulness of early intervention, and family-centered services were related to family outcomes.

A second study (Olmsted et al., 2010) included families in Illinois and Texas who completed a Spanish version of the FOS (n = 291), comparing their reported outcomes with Hispanic (n = 486) and non-Hispanic (n = 2,363) families who completed the English version of the survey. Families who completed the Spanish version reported significantly lower outcome attainment than did Hispanic and non-Hispanic families who completed the English version. They also reported lower perceptions of the helpfulness of early intervention. The three groups did not differ with regard to perceptions of family-centered practices. Factor analysis revealed that the two clusters identified by Raspa et al. (2010) were also found in the Spanish version of the survey. Analysis within the Spanish-language group replicated the Raspa et al. finding that family-centered practices were significantly related to family outcomes.

Although the FOS has undergone considerable development and evaluation and appears to have some degree of social validity, as evidenced by its use in several states and early intervention programs, users have raised several issues. First, the survey format uses a 7-point scale, with each item using unique descriptors for ratings of 1, 3, 5, and 7; that is, no descriptors are provided for ratings of 2, 4, and 6, to allow for in-between ratings. This format has been problematic for some families. It has created a perception of a longer and more complicated survey, because families are required to read each question and set of response options. Raspa et al. (2010) found that parents were less likely to choose the intermediate ratings. Second, the items may not have adequately represented the breadth of concepts that should be assessed regarding family outcomes. Several states suggested that more items might make the scale more useful for program improvement purposes, such as

identifying needs for staff development or changes in policy or practice. Finally, states and federal officials raised questions regarding whether a single item was a valid way to report data to OSEP for each of the three required indicators of helpfulness.

To address these issues, ECO investigators partnered with the Texas Department of Assistive and Rehabilitative Services and the Illinois Department of Human Services to (a) create a new format that would be easier for parents to understand, (b) revise and expand the survey items to provide more information for states to use in planning for program improvement, and (c) evaluate the psychometric properties of the expanded instrument. This article is divided into two sections. The first describes the process of generating and selecting new items; the second presents data evaluating the psychometric properties of the scale and describes changes in the instrument as a result of pilot testing. We conclude with a set of observations about family outcome assessment more broadly and recommendations for further study.

Study 1: Content and Item Development

The purpose of Study 1 was to create a new bank of potential survey items to address concerns about the limited content and usefulness of the survey for federal reporting and program improvement. We used nominations from key stakeholders and a systematic review process to (a) identify concepts to cover and (b) develop items to assess those concepts. Because the five family outcomes had already been identified through an extensive national process and the three helpfulness indicators are mandated by OSEP, we limited our work to these two frameworks, recognizing that other outcomes and other dimensions of program helpfulness could have been included.

Method

Participants. Study 1 participants included two special education researchers, one survey methodologist, two doctoral students in early childhood special education, two state Part C early intervention evaluation coordinators, and a local early intervention program coordinator. These participants represented the core study team (and coauthors of this article) who assimilated the information gathered from stakeholders: Part C early intervention evaluation or program coordinators from 19 other states, three members of the ECO advisory board, and numerous ECO staff members.

Procedure. In general, we followed recommended practices in developing or revising the content of an assessment instrument (DeVellis, 2003; Pedhazur & Schmelkin, 1991), with the goal of generating a pool of concepts that the scale should measure. To begin, we reviewed the literature on family outcomes and instruments used to assess program practices (see sources reviewed by Bailey, Bruder, et al., 2006, in generating the original five family outcomes). Information was summarized from more than 20 articles, reports, and instruments. For each family outcome and each helpfulness indicator, we listed concepts recommended in the literature that related to family outcomes. Additional concepts not related to the current items were also listed.

Equally important was the need to gather input from stakeholders and experts in the field. We developed a nomination form that we sent to early intervention officials from 19 states who were using the original version of the scale, three members of the ECO advisory board who were experts in families and family assessment, and ECO staff. The form asked respondents to list up to 10 concepts or ideas that they believed would be critical to measure when assessing the five outcomes and three helpfulness indicators. Respondents were instructed not to worry about writing specific survey questions but instead to focus on the concepts or ideas that best addressed the outcome or indicator. Responses could be provided as phrases, concepts, or complete sentences. Twenty completed forms were received with a total of 1,164 nominations. The number of nominations for each outcome ranged from 130 to 183, and the number of nominations for each OSEP helpfulness indicator ranged from 119 to 125. Nominations were compiled into an Excel spreadsheet and organized by the five outcomes and three helpfulness indicators. All nominations were reviewed and duplicates of the same idea or concept deleted.

Next, we used a modified Q-sort procedure to further refine the list of nominations. Similar to factor analysis, the Q-sort uses subjective decisions made by raters instead of statistical analyses to reduce items into subgroups of ideas or concepts (Carr, 1992). This procedure helped reduce the 1,164 nominations to a number of common themes within each of the five family outcomes and three helpfulness indicators. Each nominated concept was printed on a separate piece of paper. Working with each outcome and each indicator, two pairs of raters (one doctoral student, one researcher) independently sorted the concepts into related piles based on a common theme. For example, if five concepts were about knowing one's rights, three about exercising one's rights, and seven about how to access information about rights, three piles were created. Concepts that did not appear to be related to the outcomes or indicators were set aside. When appropriate, raters moved concepts to achieve more appropriate grouping, such as those submitted under one outcome or indicator but clearly more related to another. Once the piles were created, the groups independently labeled each pile and ranked the concepts in order of perceived importance and relevance (from most important on top to least important on the bottom). Together, both groups then discussed the sorted results to resolve differences and arrive at a consensus on the piles of concepts and their labels.

Results

We organized the final list of concepts by each outcome and helpfulness indicator, which provided the starting point for discussions with our state colleagues on the list of expanded concepts. From the list of the concepts that resulted from the Q-sort procedure, two or three examples were included under each to illustrate the main idea. Between 5 and 10 concepts were listed for each family outcome and program indicator. A draft of the concepts was shared with our state colleagues, followed by a series of conference calls. Discussions focused on the categories of concepts and whether they should be consolidated further or separated and expanded.

Once consensus was reached on the concepts, we created 27 draft family outcome items and 24 draft helpfulness indicator items. Additional conference calls were conducted with our state colleagues, who provided additional written feedback. Another round of revisions

was made that included modifying item wording, reorganizing the order of items, adding two items to the family outcomes section, and deleting one of the helpfulness indicator items. The list of items was distributed to the same group of stakeholders and experts for additional input. Based on this feedback and subsequent discussions with state officials in the Texas and Illinois early intervention programs, a field-test version of the revised survey was finalized, with 29 family outcome items and 23 helpfulness indicator items.

Next, we considered various response options for the outcome and indicator items. Several examples were created using 4- and 5-point scales, and each option was discussed by the project team. After consideration of each option, a 5-point scale was selected for the outcome and indicator items. The general instructions for the outcome items directed families to rate "which response best describes your family right now: *not at all* (1), *a little* (2), *somewhat* (3), *almost* (4), or *completely* (5)." The instructions for the helpfulness indicator items asked families to rate "how helpful early intervention has been to you and your family over the past year: *not at all helpful* (1), *a little helpful* (2), *somewhat helpful* (3), *very helpful* (4), or *extremely helpful* (5)."

Once the expanded survey was finalized, cognitive testing interviews were conducted with two parents of children in early intervention to assess whether there were any issues with question wording, instructions, formatting, or other factors that could cause survey error (Willis, 2004, 2005). Parents were asked to complete each page of the survey and then prompted with a number of questions to assess their understanding of the information presented. Parents were also asked to provide additional feedback on other issues or concerns they had about the survey. Each interview took approximately an hour.

The findings from the testing were generally positive. Parents indicated that the instructions were easy to read and understand. Both participants indicated that they believed that the questions matched the outcomes or indicators of interest and that an adequate number of response options were provided. The formatting of the scales was effective and easy to follow; in particular, parents remarked that having the questions presented in blocks showing the five outcomes and three indicators was helpful in setting the context for the questions being asked. The participants also suggested that telling parents why the questions were being asked and how the information would be used would be helpful in addressing any concerns that parents might have about the data being collected.

Following the cognitive interviews, minor modifications were made to address issues with the instructions. A brief statement was added explaining why the data were being collected. An explanation was added to indicate that references to "we" in the outcomes scale and "you" in the indicator scale referred to the survey respondents and their families. The scale was formatted to display all family outcome questions on one page and all helpfulness indicators on a second page.

Study 2: Pilot Testing and Instrument Refinement

Study 2 involved a pilot test to evaluate the psychometric properties of the instrument. The primary goal was to determine the extent to which items were statistically associated with each outcome and helpfulness indicator, to eliminate items with weak associations, and to determine whether the final set of items represented a statistically meaningful summary of each family outcome and helpfulness indicator.

Method

Participants. A convenience sample of 265 parents with a child currently enrolled in early intervention completed the revised version of the survey. Sixty-nine percent of the families were from Texas and 31% from Illinois. The majority of respondents were mothers (85%); however, fathers (5%), grandmothers (3%), and other family members (7%) participated in the survey. Just over a quarter (26%) of the respondents had a high school degree or less. Another 36% had a trade or technical school certificate, some college, or a 2-year college degree. Twenty percent had a 4-year college degree and 18% a graduate or professional degree. Family members reported a range of incomes: less than $25,000 (38%), $25,000 to $50,000 (18%), $50,001 to $75,000 (18%), $75,001 to $100,000 (14%), and over $100,000 (13%).

Families were asked to provide information on their children who were receiving early intervention services. The majority of children were male (66%). Most children were Caucasian (65%) with smaller percentages of African American (9%), Asian (4%), and other ethnicities (22%). About 44% reported that their children were of Hispanic origin. Children began receiving early intervention services before 12 months of age (43%), between 12 and 24 months of age (33%), or after 24 months of age (24%). On average, children were 25.3 months of age (SD = 8.5) at the time of the survey.

Procedures. Families were sent a cover letter by early intervention program officials in their states, inviting them to participate in a pilot study using the revised survey. The cover letter explained the purpose of the survey, provided contact information should parents have questions, and gave instructions on how to complete the survey. Families were offered the choice of completing the survey online or on paper. Paper-and-pencil surveys were entered online by early intervention staff in each state. Data collection was completed over a 4-week period.

Data analysis. Analyses were conducted using the SAS 9.2 (SAS Institute Inc., 2010) and MPlus (Muthén & Muthén, 1998-2010). Descriptive statistics include frequency distributions, means, standard deviations, as well as skewness and kurtosis for the outcome and indicator items. A number of analyses were conducted to evaluate the psychometric properties of the items, subscales, and overall scales. First, we computed Cronbach's alpha to examine interitem correlation within each subscale. Although a commonly used measure of internal consistency, Cronbach's alpha is sensitive to the number of scale items, and the interitem correlation may be increased by simply adding the number of items within each subscale (Embretson, 1999). To provide a more rigorous examination of each scale, we performed a series of confirmatory factor analyses (CFAs) for each of the five subscales in the family outcomes section and the three subscales in the program helpfulness section. We chose CFA because the underlying model structure for each distinct factor and subscale was developed by subject matter experts; hence, our analyses are hypothesis driven. Furthermore, we used two approaches to model testing within the CFA framework. First, each subscale was separately tested because policy makers may wish to implement a smaller subset rather than the full range of scales. However, we repeated our analyses assuming that each subscale was part of a second-order factor, which allowed us to estimate their interscale

correlations as an approach to understanding their interdependence. For the CFA, we used the maximum likelihood estimator, which is appropriate for Likert-type distributions (Bollen & Long, 1993). In the CFA, three initial criteria were used to determine whether a scale had good psychometric properties (Hoyle, 1995; MacCallum, 2003): (a) a nonsignificant chi-square ($p > .05$), (b) a root mean square error of approximation (RMSEA) of .04 or below, and (c) a value of .90 or greater on the Tucker–Lewis index (TLI) and comparative fit index (CFI). In instances in which the subscale exhibited poor psychometric properties, individual item loadings using a two-parameter (item discrimination and item difficulty) item response theory (IRT) model for polytomously distributed response categories (i.e., *not at all*, *a little*, *somewhat*, *almost*, and *completely*; Embretson & Reise, 2000). The CFA allows us to examine the correlations among the measured items along a single latent trait for each scale. The complementary IRT model allows for inspection of the relation between the latent trait and the item's properties and their joint influence on the probability of endorsing a specific response option within a single item. Put differently, CFA enables an understanding of how the latent trait may be improved by adding or removing items, whereas the IRT model allows us to not only understand how characteristics of each item are related to the underlying latent trait but also consider the distribution of response options within each item. As each technique makes much different, yet complementary, assumptions about the relation between the latent trait and individual items, joint inspection of their results permits a more complete picture of the psychometric properties. This range of techniques is critical when external validation samples are not available. For the IRT models, we inspected the discrimination and difficulty parameters using a graded-response option that considers the categorical nature of the Likert scale (Samejima, 1969). We plotted the item characteristic curves, the item's endorsement probability (difficulty), and its ability to differentiate (discrimination) respondents with better or worse outcomes along the latent trait of each subscale. For brevity, the IRT results were consistent with the findings from the CFA, and we referred to the results of the IRT analyses when differences suggested removal of items.

After inspection of the IRT models, subject matter experts considered the value of each candidate item, and some items that exhibited poorer fit were retained because of their policy implications. The CFA was then re-estimated after problematic items were dropped. In the final step, the overall factor structures of the outcome scale and indicator scale were examined to identify whether a single-factor solution was warranted rather than a second-order factor that included each subscale simultaneously. All significance levels reported herein represent two-tailed tests at $\alpha < .05$.

Results

Descriptive statistics. Table 1 provides the response distribution for the 29 items on the family outcomes portion of the survey, and Table 2 indicates the distribution in terms of skew and kurtosis. Overall, respondents indicated that they had achieved a high percentage of family outcomes. Each item was positively skewed, with the majority of respondents indicating that they almost or completely achieved the outcome statement. The items with the highest mean scores were Item 4 ("We are able to tell when our child is making progress"), Item 9 ("We are comfortable talking with service providers about our child's needs"), Item 10 ("We are comfortable asking for services and supports that our child and family

Table 1
Mean, Standard Deviation, and Response Distribution for the Family Outcome Items

		M (SD)	Not at All	A Little	Somewhat	Almost	Completely
Understanding Strengths							
1	We know next steps in development	4.38 (0.79)	0%	2%	13%	30%	55%
2	We understand child's strengths	4.63 (0.67)	0	2	6	21	71
3	We understand child's needs	4.61 (0.66)	0	1	7	22	70
4	We are able to tell child's progress	4.77 (0.58)	0	2	2	14	82
Knowing Rights							
5	We use available services	4.51 (0.78)	0	3	10	21	66
6	We know our rights	4.42 (0.92)	2	3	12	19	64
7	We know who to contact with questions	4.69 (0.63)	< 1	< 1	6	17	77
8	We know options after leaving program	4.07 (1.22)	6	7	13	22	52
9	We are comfortable talking to provider	4.83 (0.52)	0	1	4	7	88
10	We are comfortable asking for services	4.76 (0.58)	0	1	5	10	83
11	We are comfortable making decisions	4.84 (0.48)	< 1	0	2	10	88
12	We feel like team members	4.83 (0.51)	0	1	2	9	88
Helping Child Develop and Learn							
13	We help child get along with others	4.44 (0.74)	0	2	11	29	58
14	We help child learn new skills	4.60 (0.68)	0	2	5	23	70
15	We help child take care of needs	4.62 (0.72)	1	1	6	20	72
16	We work on goals during routines	4.59 (0.70)	0	2	7	22	69
17	We change routines to meet needs	4.54 (0.69)	0	2	7	27	64
18	We select materials to help child learn	4.59 (0.72)	0	2	7	20	71
Having Support Systems							
19	We talk with friends/family about needs	4.61 (0.82)	2	2	5	16	75
20	We have friends/family who listen	4.64 (0.83)	1	4	4	11	80
21	We talk with families with similar child	4.02 (1.35)	9	7	15	11	58
22	We have friends/family to rely on	4.42 (0.99)	2	4	11	16	67
23	I take care of my own needs	4.32 (1.01)	2	4	13	19	61
Accessing the Community							
24	Participate in social/recreational/religious acts	4.39 (1.03)	4	4	9	18	66
25	We do things together as a family	4.53 (0.90)	2	3	8	14	73
26	Medical and dental needs met	4.68 (0.73)	7	2	4	14	79
27	Child care needs met	4.61 (0.81)	1	2	8	12	76
28	Transportation needs met	4.72 (0.72)	1	2	4	10	83
29	Food, clothing, housing needs met	4.72 (0.68)	< 1	3	4	12	82

need"), Item 11 ("We are comfortable helping make decisions about our child's services and supports"), and Item 12 ("We feel like important members of the team"). The items with the lowest mean scores were Item 1 ("We know the next steps for our child's growth and learning"), Item 8 ("We know what options are available when our child leaves the program"), Item 21 ("We are able to talk with other families who have a child with similar needs"), Item 23 ("I am able to take care of my own needs and do things I enjoy"), and Item 24 ("Our child is able to participate in social, recreational, or religious activities that we want").

Table 3 presents the response distribution for the 22 helpfulness indicator items, and Table 4 presents the distributional statistics. In general, families reported a high level of helpfulness of early intervention programs. The majority of families responded that their program was either very helpful or extremely helpful over the past year. Although the responses were positively skewed, fewer respondents selected the highest response level (*extremely helpful*) for the helpfulness items compared to the highest response level (*completely*) for the family outcome items. The items with the highest means were Item 6 ("Making you feel comfortable sharing your concerns"), Item 9 ("Scheduling meetings and services at times that work for you"), Item 13 ("Making you feel like an important member of the team"), and Item 14 ("Developing a good relationship with you and your family"). The items with the lowest means were Item 2 ("Giving you useful information about your rights related to your child's special needs"), Item 4 ("Giving you information about available options when your child leaves the program"), and Item 15 ("Giving you useful information about how to help your child get along with others").

Psychometric analyses of the family outcome items. In general, the expanded set of family outcome items showed good psychometric properties, as shown in Table 2. However, a number of items were identified as either not contributing to the scale or potentially providing redundant information. As described below, these items were removed and scales re-analyzed, which improved internal consistency.

Outcome 1 (Understanding Your Child's Strengths, Needs, and Abilities) originally consisted of four questions (Items 1–4). Cronbach's alpha for this subscale was .73, and the CFA model exhibited good psychometric properties ($\chi^2 = 2.42$, 2 $df = 2$, $p = .296$, TLI = .99, CFI = .99, RMSEA = .028). Therefore, no modifications were made to this subscale.

Outcome 2 (Knowing Your Rights and Advocating for Your Child) comprised eight questions (Items 5–12). Cronbach's alpha for this subscale was .82, but the CFA model exhibited poor psychometric properties ($\chi^2 = 212.60$, $df = 20$, $p < .001$, TLI = .68, CFI = .77, RMSEA = .193). Items 9, 11, and 12 had the lowest factor loadings. After those items were dropped, the alpha for this subscale was .78, and the model exhibited good psychometric properties ($\chi^2 = 6.58$, $df = 5$, $p = .252$, TLI = .99, CFI = .99, RMSEA = .035).

Outcome 3 (Helping Your Child Develop and Learn) contained six questions (Items 13–18). Cronbach's alpha for this subscale was .89, but the CFA model exhibited poor psychometric properties ($\chi^2 = 180.51$, $df = 9$, $p < .001$, TLI = .71, CFI = .83, RMSEA = .27). After Items 17 and 18 were dropped because of low standardized loadings, the subscale exhibited good fit. Cronbach's alpha was .87, and the fit indices were acceptable ($\chi^2 = 8.80$, $df = 2$, $p = .01$, TLI = .96, CFI = .99, RMSEA = .12).

Outcome 4 (Having Support Systems) consisted of five questions (Items 19–23). The alpha for this scale was .78, and the CFA model exhibited reasonable psychometric properties

Table 2
Psychometric Analyses for Family Outcome Items

Scale	Skewness	Kurtosis	Full Scale	Reduced
1: Understanding Strengths				
Item 1	–1.02	0.16	.687	N/E
Item 2	–1.85	3.12	.765	N/E
Item 3	–1.58	1.83	.642	N/E
Item 4	–2.95	9.52	.484	N/E
α			.73	N/E
Factor variance			.29	
χ^2 *(df)*			2.42 (2)	
CFI			.998	
TLI			.994	
RMSEA			.028	
Interscale (second-order factor) correlation with:				
Scale 2				.572*
Scale 3				.622*
Scale 4				.569*
Scale 5				.467*
2: Knowing Rights				
Item 5	–1.53	1.52	.653	.801
Item 6	–1.64	2.25	.673	.874
Item 7	–2.32	6.18	.537	.558
Item 8	–1.19	0.37	.541	.621
Item 9	–3.30	10.98	.731	N/E
Item 10	–2.57	6.19	.701	.478
Item 11	–3.90	19.98	.675	N/E
Item 12	–3.51	13.34	.729	N/E
α			.82	.78
Factor variance			.26	.39
χ^2 *(df)*			212.60 (20)*	6.58 (10)
CFI			.777	.996
TLI			.688	.992
RMSEA			.193	.035
Interscale (second-order factor) correlation with:				
Scale 1				.572*
Scale 3				.486*
Scale 4				.446*
Scale 5				.278*
3: Helping Child Develop and Learn				
Item 13	–1.15	0.57	.821	.795
Item 14	–1.82	3.12	.802	.935
Item 15	–2.24	5.81	.734	.767
Item 16	–1.15	0.57	.879	.723
Item 17	–1.48	1.79	.734	N/E
Item 18	–1.82	2.77	.741	N/E
α			.89	.87
Factor variance			.29	.32
χ^2 *(df)*			180.51 (9)*	8.79 (2)*
CFI			.827	.998
TLI			.711	.964
RMSEA			.272	.115
Interscale (second-order factor) correlation with:				
Scale 1				.622*
Scale 2				.486*
Scale 4				.633*
Scale 5				.529*

(continued)

Table 2 (continued)

Scale	Skewness	Kurtosis	Full Scale	Reduced
4: Having Support Systems				
Item 19	–2.51	6.42	.759	N/E
Item 20	–2.57	6.18	.875	N/E
Item 21	–1.11	–0.11	.447	N/E
Item 22	–1.76	2.45	.736	N/E
Item 23	–1.47	1.48	.554	N/E
α			.78	
Factor variance			.38	
χ^2 *(df)*			26.30 (5)*	
CFI			.952	
TLI			.904	
RMSEA			.121	
Interscale (second-order factor) correlation with:				
Scale 1				.569*
Scale 2				.446*
Scale 3				.633*
Scale 5				.685*
5: Accessing the Community				
Item 24	–1.82	2.71	.844	N/E
Item 25	–2.08	3.88	.851	N/E
Item 26	–2.74	7.97	.432	—[a]
Item 27	–2.28	5.08	.555	N/E
Item 28	–3.01	9.33	.446	—[a]
Item 29	–2.83	8.26	.498	N/E
α			.91	
Factor variance			.74	
χ^2 *(df)*			175.91 (9)*	
CFI			.711	
TLI			.517	
RMSEA			.268	
Interscale (second-order factor) correlation with:				
Scale 1				.467*
Scale 2				.278*
Scale 3				.529*
Scale 4				.685*

Note. N/E = item not estimated in reduced model; α = Cronbach α for the scale; factor = the standardized regression coefficients derived from the confirmatory factor analysis, representing the variance of the latent factor; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation.
a. Item retained in substantive review.
*$p < .05$.

($\chi^2 = 26.30$, $df = 5$, $p < .001$, TLI = .90, CFI = .95, RMSEA = .12), and thus no changes were made.

Outcome 5 (Accessing the Community) comprised six questions (Items 24–29). The alpha for this subscale was .91, and the CFA model exhibited good psychometric properties ($\chi^2 = 175.91$, $df = 9$, $p < .001$, TLI = .52, CFI = .71, RMSEA = .26). Items 26 and 28 exhibited relatively poor item fit based on the IRT analysis. These questions were retained, however, because the overall model was acceptable and these were the only items dealing with how well medical, dental, and transportation needs were being met—content deemed highly important by stakeholders.

Table 3
Mean, Standard Deviation, and Response Distribution for the Helpfulness Indicator Items

		M (SD)	Not at All	A Little	Somewhat	Almost	Completely
Knowing Your Rights							
1	Information about services	4.54 (0.70)	0%	2%	4%	30%	63%
2	Information about rights	4.43 (0.86)	1	3	7	28	61
3	Information about who to contact	4.56 (0.73)	1	2	5	26	66
4	Information about transition	4.25 (1.09)	4	7	6	27	56
5	Explained rights	4.54 (0.77)	1	2	5	25	66
6	Felt comfortable sharing concerns	4.71 (0.59)	< 1	1	2	20	76
Communicating Child's Needs							
7	Information about child's needs	4.57 (0.70)	1	1	5	27	66
8	Listened and respected your choices	4.68 (0.37)	1	2	2	24	75
9	Scheduled convenient meetings	4.70 (0.67)	1	0	5	16	78
10	Connected you with services/people	4.68 (0.89)	1	4	8	29	57
11	Talked with you about child's needs	4.54 (0.69)	< 1	1	7	29	64
12	Talked with you about priorities	4.59 (0.65)	< 1	1	4	29	66
13	Felt like team member	4.70 (0.59)	0	1	4	19	76
14	Developed good relationship	4.71 (0.62)	< 1	1	4	18	77
Helping Child Develop and Learn							
15	Information about how to get along	4.30 (0.96)	3	2	11	30	54
16	Information about learning new skills	4.54 (0.74)	< 1	3	6	27	65
17	Information about taking care of needs	4.52 (0.78)	1	2	6	26	65
18	Identifying what you do to help learn	4.55 (0.76)	1	3	3	27	67
19	Sharing ideas about routines	4.53 (0.75)	< 1	3	5	28	64
20	Showing how to work with child	4.51 (0.82)	1	3	6	21	69
21	Chance to practice new skills	4.55 (0.80)	1	3	5	22	69
22	Helping to identify child's progress	4.58 (0.77)	1	2	4	23	70

Table 4
Psychometric Analyses for Helpfulness Indicator Items

Scale	Skewness	Kurtosis	Full Scale	Reduced
1: Knowing Your Rights				
Item 1	–1.62	2.74	.849	N/E
Item 2	–1.74	3.03	.924	.927
Item 3	–2.03	5.04	.859	.829
Item 4	–1.55	1.64	.722	.742
Item 5	–1.98	4.38	.902	.921
Item 6	–2.61	9.54	.694	N/E
α			.73	.90
Factor variance			.34	.63
χ^2 *(df)*			74.53 (9)*	2.22 (2)
CFI			.948	.99
TLI			.814	.99
RMSEA			.168	.021
Interscale (second-order factor) correlation with:				
Scale 2				.796*
Scale 3				.686*
2: Communicating Child's Needs				
Item 7	–2.03	5.49	.787	.801
Item 8	–2.77	9.54	.886	.837
Item 9	–2.89	10.47	.751	N/E
Item 10	–1.58	2.31	.607	.628
Item 11	–1.62	3.16	.773	.836
Item 12	–1.83	4.53	.874	.911
Item 13	–2.07	4.16	.889	N/E
Item 14	–2.53	7.81	.882	.85
α			.92	.91
Factor variance			.31	.31
χ^2 *(df)*			175.63 (20)*	74.24 (9)*
CFI			.911	.943
TLI			.875	.905
RMSEA			.174	.168
Interscale (second-order factor) correlation with:				
Scale 1				.796*
Scale 3				.851*
3: Helping Child Develop and Learn				
Item 15	–1.62	2.68	.721	.721
Item 16	–1.82	3.63	.894	.906
Item 17	–1.91	3.99	.874	.915
Item 18	–2.16	5.31	.869	.881
Item 19	–1.84	3.72	.855	.865
Item 20	–2.12	4.65	.866	N/E
Item 21	–2.04	4.26	.821	N/E
Item 22	–2.32	6.21	.841	.773
α			.95	.94
Factor variance			.46	.46
χ^2 *(df)*			238.59 (20)*	27.74 (9)*
CFI			.894	.986
TLI			.852	.976
RMSEA			.206	.091
Interscale (second-order factor) correlation with:				
Scale 1				.686*
Scale 2				.851*

Note. N/E = item not estimated in reduced model; α = Cronbach α for the scale; factor = the standardized regression coefficients derived from the confirmatory factor analysis, representing the variance of the latent factor; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation.
*$p < .05$.

In addition to examining the five subscales individually, we evaluated whether the outcome items functioned better as a single scale or as a set of five subscales aligned with the five outcomes. This analysis incorporated the subset of items that were retained after poor-performing items were dropped, as described above. The alpha for the single-factor scale was .90, but the CFA model exhibited poor psychometric properties (χ^2 = 1,487, *df* = 252, $p < .001$, TLI = .53, CFI = .57, RMSEA = .14). A second-order factor with five subscales was fit, and this significantly improved model fit (χ^2 = 752.51, *df* = 243, $p < .001$, TLI = .80, CF I = .82, RMSEA = .09). These results suggest that the scale functions better as a set of five individual subscales rather than as a single, overall scale of family outcomes.

Psychometric analyses of the helpfulness indicator items. In general, the expanded set of items assessing program helpfulness showed adequate psychometric properties, as shown in Table 4. A number of items were identified as either not contributing to the scale or potentially providing redundant information. Therefore, these items were removed and the scales re-analyzed to improve internal consistency.

Indicator 1 (Knowing Your Rights) comprised six questions (Items 1–6). Cronbach's alpha for this subscale was .73, and the CFA model exhibited good psychometric properties (χ^2 = 74.53, *df* = 9, $p < .001$, TLI = .81, CFI = .95, RMSEA = .168); however, the IRT analyses indicated that dropping Items 1 and 6 would improve model fit. After these two items were dropped, the model exhibited excellent fit (α = .90, χ^2 = 2.22, *df* = 2, p = .329, TLI = .99, CFI = .99, RMSEA = .021).

Indicator 2 (Communicating Your Child's Needs) consisted of eight questions (Items 7–14). Cronbach's alpha for this subscale was .92, and the CFA model exhibited acceptable psychometric properties (χ^2 = 175.63, *df* = 20, $p < .001$, TLI = .87, CFI = .91, RMSEA = .174); however, the IRT analyses showed significant lack of discrimination for Items 9 and 13. After these two items were dropped, the psychometric properties improved (α = .91, χ^2 = 74.24, *df* = 9, $p < .001$, TLI = .91, CFI = .94, RMSEA = .168).

Indicator 3 (Helping Your Child Develop and Learn) contained eight questions (Items 15–22). The alpha for this subscale was .95, although the CFA model exhibited poor psychometric properties (χ^2 = 238.59, *df* = 20, $p < .001$, TLI = .85, CFI = .89, RMSEA = .206). After Items 20 and 21 were dropped, the subscale exhibited good fit (α = .94, χ^2 = 27.74, *df* = 9, $p < .001$, TLI = .98, CFI = .99, RMSEA = .09).

Finally, we evaluated whether all the helpfulness questions functioned better as a single scale or three subscales. This analysis was conducted on the subset of questions that were retained after each subscale was analyzed. In general, the findings showed that the helpfulness indicator items are heterogeneous and do not hold together well in a single factor or in three subscales, the latter of which appeared to function better as individual scales but were not substantially better than a single factor. The overall alpha was .96, and the single factor model exhibited poor fit (χ^2 = 2,666.80, *df* = 213, $p < .001$, TLI = .56, CFI = .59, RMSEA = .211). The second-order factor with three subscales slightly improved model fit but not substantially (χ^2 = 2,090.47, *df* = 210, $p < .001$, TLI = .66, CFI = .69, RMSEA = .186).

Correspondence with the single-item helpfulness scale. A subset of families (n = 44) completed the three items on the original version of the FOS that had been designed for states to use in reporting to OSEP. Pearson's correlation coefficient indicated that each final set of helpfulness items was significantly related to the single item on the original scale.

For Indicator 1 (Knowing Your Rights), the correlation was .72; Indicator 2 (Communicating Your Child's Needs), .83; and Indicator 3 (Helping Your Child Develop and Learn), .78.

Discussion

We have described herein the steps followed, data collected, and decisions made in revising and expanding the FOS. There were three goals of this work: (a) to make the FOS easier for parents to complete; (b) to provide more information that states could use in making decisions about program improvement; and (c) to evaluate the psychometric properties of the revised scale.

Using multiple sources of input from subject matter experts, we generated additional items of potential use in the assessment of family outcomes and the helpfulness of early intervention. The items were then field-tested with a sample of families in Illinois and Texas. We started with a full scale of 29 items for the family outcomes scale and 22 for the helpfulness scale. Psychometric analyses suggested that some items did not contribute well to the overall estimation of the underlying latent scales. Most of these items were dropped from the survey, resulting in a scale with 24 items assessing the extent to which family outcomes are attained and 17 items assessing the extent to which families perceive early intervention to have been helpful. The revised instrument has been posted on the ECO Center website (http://www.the-eco-center.org). The website includes translations into Arabic, Chinese, Croatian, Hmong, Khmer, Korean, Lao, Russian, Somali, Spanish, and Vietnamese; other translations will be posted as they become available.

The revised instrument has several potential uses. For purposes of research and evaluation, data from the outcomes portion of the scale could be used as evidence of whether a program, region, or state is resulting in perceived benefit to families. Summary scores could be generated for each of the five outcome areas and a total score used to reflect overall levels of outcome attainment. Analyses revealed, however, that the scale appears to function best as a group of independent measures rather than as a single family outcome or indicator of helpfulness. The helpfulness scales could be used to report data to OSEP. For reporting purposes, we recommend that an average score of 4 or higher be used for the items related to one of the three indicators, to determine whether the indicator has been met or not. The indicator items could also be used in research or evaluation efforts to reflect family perceptions of intervention helpfulness. However, because these items were generated specifically for the federal reporting requirements, other instruments might provide a more comprehensive assessment of family satisfaction or family-centered practices. Finally, the additional items were added in part to provide information that could be used for program improvement purposes. For this purpose, it would be useful to look at the distribution of individual items to identify specific areas in which outcomes may not be achieved or in which helpfulness is perceived as being low.

The study has several limitations. First, we structured the nomination process around the five family outcomes and three helpfulness indicators, thus limiting the content of both the nominations and the resulting scale. Other potentially important family outcomes or dimensions of services and helpfulness of programs were therefore not considered. Second, the nominators and the participants in the pilot study may not be representative of all stakeholders

or consumers of early intervention services. A larger and more diverse sample may have resulted in additional or different suggestions for content. Families who are dissatisfied with early intervention or those who dropped out of early intervention are not represented here, and their opinions may have changed the study findings and recommendations. Third, we did not have sufficient power to conduct tests of measurement invariance to identify whether there were differences in mode of administration (i.e., paper versus online). Future studies should be designed to evaluate the instrument in various modes of administration to determine if differential response patterns occur. Finally, we did not address the issues of validity. The extent to which the scale provides a true and objective assessment of the real outcomes experienced by families or the real quality of services has not yet been demonstrated.

Despite these limitations, the revised survey stands as an easier, more comprehensive, and potentially more useful instrument for the assessment of family outcomes and the helpfulness of early intervention programs. Further research is needed to determine the scale's ultimate utility as part of an overall evaluation effort and to better understand the relationships among outcomes and services. Although one could argue that parents are the ultimate stakeholders and that their reports of outcomes, helpfulness, and benefit should be sufficiently valid, it would be useful to conduct research comparing parent self-reports with other indices of benefit and helpfulness.

References

Bailey, D. B. (2001). Evaluating parent involvement and family support in early intervention and preschool programs: Three levels of accountability. *Journal of Early Intervention, 24*, 1-14.

Bailey, D. B., Bruder, M., Hebbeler, K., Carta, J., de Fosset, M., Greenwood, C., . . . Barton, L. (2006). Recommended outcomes for families of young children with disabilities. *Journal of Early Intervention, 28*, 227-251.

Bailey, D. B., Buysse, V., Edmondson, R., & Smith, T. M. (1992). Creating family-centered services in early intervention: Perceptions of professionals in four states. *Exceptional Children, 58,* 298-309.

Bailey, D. B., Hebbeler, K., & Bruder, M. B. (2006). *Family Outcomes Survey*. Retrieved April 19, 2010, from http://www.fpg.unc.edu/~eco/assets/pdfs/FOS-PartC_11-16-06.pdf

Bailey, D. B., Hebbeler, K., Olmsted, M., Raspa, M., & Bruder, M. B. (2008). Measuring family outcomes: Considerations for large-scale data collection in early intervention. *Infants & Young Children, 21,* 194-206.

Bailey, D. B., Hebbeler, K., Spiker, D., Scarborough, A., & Mallik, S. (2004). First experiences with early intervention: A national perspective. *Pediatrics, 113,* 887-896.

Bailey, D. B., Simeonsson, R. J., Winton, P. J., Huntington G. S., Comfort, M., Isbell, P., . . . Helm, J. M. (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.

Bollen, K. A., & Long, J. S. (Eds.). 1993. *Testing structural equation models.* Newbury Park, CA: Sage.

Carr, S. C. (1992). A primer on the use of the Q-technique factor analysis. *Measurement and Evaluation in Counseling and Development*, *25*(3), 133-138.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Dunst, C. J. (1985). Rethinking early intervention. *Analysis and Intervention in Developmental Disabilities, 5,* 165-201.

Embretson, S. E. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika, 64,* 407-433.

Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Hebbeler, K. M., & Barton, L. R. (2007). The need for data on child and family outcomes at the federal and state levels. In E. Horn, C. M Peterson, & L. Fox (Eds.), *Young Exceptional Children Monograph Series No. 9: Linking curriculum to child and family outcomes* (pp. 1-15). Longmont, CO: Sopris West.

Hoyle, R. H. (Ed.). (1995). *Structural equation modeling: Concepts, issues, and applications.* Thousand Oaks, CA: Sage.

Iversen, M. D., Shimmel, J. P., Ciacera, S. L., & Prabhakar, M. (2003). Creating a family-centered approach to early intervention services: Perceptions of parents and professionals. *Pediatric Physical Therapy, 15,* 23-32.

Kahn, L. (2009, June). *National update from NECTAC/ECO*. Paper presented at the Measuring Child and Family Outcomes Conference, Bethesda, MD.

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research, 38*, 113-139.

McWilliam, R. A., Lang, L., Vandiviere, P., Angell, R., Collins, L., & Underdown, G. (1995). Satisfaction and struggles: Family perceptions of early intervention services. *Journal of Early Intervention, 19,* 43-60.

Muthén, L. K., & Muthén, B. O. (1998-2010). Mplus user's guide (6th ed.). Los Angeles, CA: Muthén & Muthén.

Olmsted, M. G., Bailey, D. B., Raspa, M., Nelson, R., Robinson, N., Simpson, M. E., & Guillen, C. (2010). Outcomes reported by Spanish-speaking families in early intervention. *Topics in Early Childhood Special Education, 30*, 46-55.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hillside, NJ: Erlbaum.

Raspa, M., Bailey, D. B., Nelson, R., Robinson, N., Simpson, M. E., Gillian, C., . . . Houts, R. (2010). Measuring family outcomes in early intervention: Findings from a large-scale assessment. *Exceptional Children, 76,* 496-510.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34*, 100-114.

SAS Institute Inc. (2010). SAS 9.2 language reference: Concepts (2nd ed.). Cary, NC: Author.

Turnbull, A. P., Summers, J. A., Turnbull, R., Brotherson, M. J., Winton, P., Roberts, R., & Stroup-Rentier, V. (2007). Family supports and services in early intervention: A bold vision. *Journal of Early Intervention, 29,* 187-206.

Willis, G. (2004). Cognitive interviewing revisited: A useful technique, in theory? In S. Presser, J. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 23-43). Hoboken, NJ: Wiley.

Willis, G. (2005). *Cognitive interviewing: A tool for improving questionnaire design.* Thousand Oaks, CA: Sage.