

# Measuring Schools' Efforts to Partner with Parents of Children Served Under IDEA:

Scaling and Standard Setting for Accountability Reporting

August 2011

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



## **Summary**

Indicator 8 of 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 findings related to schools' facilitation of parent involvement. The Schools' Efforts to Partner with Parents Scale (SEPPS) was developed to provide states with a means to address this new reporting requirement. Items suggested by stakeholder groups were piloted with a nationally representative sample of 2,634 parents of students with disabilities ages 5-21 in six states. Rasch scaling was used to calibrate a meaningful and invariant item hierarchy. The 78 calibrated items had measurement reliabilities ranging from .94-.97. Using data from the pilot study, stakeholders established a recommended performance standard set at a meaningful point in the item hierarchy. Implications of the findings are discussed in relation to the need for rigorous metrics within state accountability systems.

## **Acknowledgement**

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

## **Correspondence**

Correspondence concerning this report should be addressed to Batya Elbaum, Department of Teaching and Learning, University of Miami, P.O. Box 248065, Coral Gables, FL 33146-2040. Email: [elbaum@miami.edu](mailto:elbaum@miami.edu).

## Introduction

Section 616 of the Individuals with Disabilities Education Act (IDEA 2004) requires states to collect data and report findings annually on a set of performance indicators described in what is called the State Performance Plan (SPP; Regional Resource Center Program, 2008a). Indicator 8 of the SPP, related to special education services provided to children ages 3-21, is “the percent of parents who report that schools facilitated their involvement as a means of improving services and results for students with disabilities.” 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) in 2002, developed and validated a tool to meet states’ need to fulfill the requirements of the SPP with regard to the parent involvement indicator.

Three components were deemed to be essential to the development of this tool. First, parents and families of children with disabilities needed to be involved in the development of the content of the instrument. Families of children with disabilities have for many years been among the strongest advocates for improved educational services for their children; additionally, the SPP requires that reporting for the parent involvement indicator be based on parents’ reports, rather than schools’ reports, of schools’ efforts to facilitate parent involvement. Accordingly, one of NCSEAM’s first steps was to establish a Parent/Family Workgroup whose members represented families of children receiving special education services, school personnel, state-level administrators of special education, special education researchers, and disability advocates. The Parent/Family Workgroup provided significant input into the instrument development process, ensuring that the resulting tool would be accessible and meaningful to parents of all educational levels and cultural backgrounds, e.g., that the items would be of a length and reading level appropriate to the intended audience and would have content that parents could readily understand.

Second, given both the high stakes and the public nature of federal accountability reporting, any data collection tool would have to possess the mathematical rigor and practical convenience of a scientifically defensible scaling methodology. The instrument developed by NCSEAM was designed to meet the most stringent measurement standards, including demonstrated scale invariance across different population subgroups, high measurement reliability ( $>.90$ ), and a common metric for the expression of measures made from any subgroup of items drawn from the larger bank of calibrated items. To accomplish these goals, NCSEAM adopted the Rasch measurement framework (Bond & Fox, 2007; Rasch, 1960; Wright & Stone, 2004) to guide evaluation of item fit, construct validation, and scale refinement.

Third, parents and other stakeholders would have to be given a voice in establishing an appropriate performance standard for reporting purposes. Thus, once data from NCSEAM's pilot study had been analyzed, a nationally representative stakeholder group was convened to develop recommendations as to the level of school facilitation of parent involvement that should be required in order to fulfill the intent of the SPP indicator. Although states have discretion to set their own performance standards and targets for future performance, drawing on recommendations of their state-level advisory committees, NCSEAM felt it was important to establish a broad, national consensus which states could then use to inform their local decisions.

In the sections that follow, methods and results are reported for the development and validation of the instrument, called the Schools' Efforts to Partner with Parents Scale (SEPPS). In the discussion, we highlight the value of rigorous measurement in educational accountability systems.

## **Method**

### **Item Development**

Items were developed on the basis of (a) published research on parent involvement, (b) survey instruments already being used in some states to gather data on parent/family perceptions and involvement, (c) stakeholder input sessions held in six states, and (d) consultation with parent organizations and professional experts. A review of previous research, along with detailed study of data from several states' existing survey instruments, informed the facilitation of stakeholder input sessions held in the spring of 2003 in six states (California, Florida, Kentucky, New Hampshire, New Mexico, and Mississippi). The number of participants in each day-long session ranged from 25 to 42, and included parents of children with disabilities, other family members, parent advocates, special education teachers and administrators, providers of related services, and community representatives.

At each meeting, participants were invited to generate items addressing a number of interrelated aspects of parents' experience with services provided under IDEA. Topics included the ways schools facilitate parent involvement, the ways in which parents participate in their child's education, and the outcomes that families achieve as a result of their participation. A draft set of 490 unique items suggested by the stakeholder groups was reviewed by representatives of the Technical Assistance Alliance for Parent Centers, also funded by OSEP, and by the NCSEAM Parent/Family Involvement Workgroup. A total of 204 items were selected for piloting; 80 of these addressed schools' facilitation of parent involvement. Items were reworded, as necessary, to make them amenable to a disagree-to-agree response format and to make them as short and easy to understand as possible. The readability of the items was assessed using the Lexile Framework for Reading (Stenner, 2001). Lexiles for the items related to schools'

facilitation of parent involvement ranged from 850 to 1020, which corresponds to the range of reading difficulty found in books written for sixth-graders.

## Procedure

Approval was obtained from the Institutional Review Board of the Louisiana State University to conduct a national field study of the items generated by stakeholder groups. State Departments of Education in six states (Florida, Georgia, Mississippi, New Hampshire, New Jersey, and New Mexico) agreed to solicit the participation of parents in their states whose children in grades pre-kindergarten through 12 were receiving special education services. Each state adhered to a customized sampling plan designed to yield a total sample across participating states whose distribution by ethnicity was very close to that of the U.S. population served under IDEA. Thus, the sampling plan for certain state samples included disproportionately high numbers of parents in groups underrepresented in other states. For instance, as shown in Table 1, the sampling plan for New Mexico specified that 40 percent of those receiving the survey should be Native Americans.

With regard to category of disability, the sampling plan was designed to under-sample parents of students with specific learning disabilities, who represent almost half of all children served under IDEA (Data Accountability Center, 2010a). This was done so as to create a more balanced distribution of responses for the anticipated testing of differences in scale properties by category of disability.

Optically scannable forms were printed and distributed to participating states. Parent recruitment strategies and modes of administration of the survey differed by state. In most cases, the survey forms were mailed to parents with a postage prepaid envelope for return of the completed survey. Several states also convened groups of parents and provided a facilitator to assist them in completing the survey. Mode and language of administration of the survey were recorded on the survey form so that it would be possible to examine whether these variables were associated with any differences in results. No personally identifying information was requested of any participants.

## Participants

Participants in the national field study consisted of 2,702 parents of a child receiving special education services. Data screening prior to analysis revealed that 11 records were missing data on the child's age, and 15 were from parents whose children were below the federally established age (i.e., 3 years) for participating in Part B special education services. These records were removed from the data set. The record for a child with a reported age of 26 was also removed, leaving a total of 2,675 records. Additionally, prior to analyzing the data for this study, a decision was made to create additional items for parents

of children receiving preschool special education services and to conduct additional validation studies for this sub-population (see limitations section, below). Thus, records of parents of children of preschool age (ages 3-4, n = 29) were removed from the data set for this study. Removal of these records yielded a data set containing a total of 2,646 records. Of these, 2,634 had valid responses to the rating scale items (12 returned surveys did not have any responses).

Table 2 provides summary statistics describing participants by state of residence, ethnicity, child's age group, child's category of disability, survey form (A or B—see below), language in which the survey was completed, and administration mode (self-administered or facilitated). For comparison purposes, the rightmost column of Table 2 displays the distribution of children ages 6-21 served under IDEA by ethnicity (Data Accountability Center, 2010b) and category of disability (Data Accountability Center, 2010a). Note that the child count data posted by OSEP on the website of the Data Accountability Center ([www.ideadata.org](http://www.ideadata.org)) are separated into a table related to children ages 3-5 and a separate table related to children 6-21. We have used the data in the table for ages 6-21 as these provide a very close, if not perfect, match to the age range of our sample. As seen in Table 2, the distribution of ethnicity in the sample is very close to that of the population served.

## Instrument

To reduce the response burden on participating parents, two alternate survey forms were created. The total pool of 204 items was divided into three groups: a common group of items which appeared on both forms, and two unique groups of items, each of which was printed on one of the two alternate forms. Form A of the survey included 158 items; form B included 160 items. Of the 80 items related to schools' facilitation of parent involvement, 24 appeared on both forms; the remaining 56 were divided equally across the two alternate forms.

An identical response scale was used for all items, consisting of six options: very strongly disagree, strongly disagree, disagree, agree, strongly agree, very strongly agree. Respondents were instructed to skip any items that they felt did not apply to them or their child.

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 provided translations in other languages, and made these translations available for use in other states that needed them.

## Results

### Scaling

Item responses from the 2,634 records containing parent responses to the NCSEAM pilot survey were analyzed through the Winsteps software program (Linacre, 2009) using a rating scale model for ordered categories (Andrich, 1978, 1996; Luo, 2005; Wright & Masters, 1982). Responses were coded from 1 (very strongly disagree) to 6 (very strongly agree). The scale was calibrated along a 0-1000 range, with the original logits rescaled by a factor of 54.105. Ceiling and floor effects were minimal (2% and 4%, respectively). Parents responding to Forms A (1,292 responses) and B (1,342 responses) provided responses to an average of about 46 school facilitation items. The data were about 59% complete; most missing data were due to using different items on the two forms.

### Scale Optimization

Examination of the response percentages in each category (some of which had fewer than 5% of all responses), the response probability curves for the original six response categories (Figure 1), and several other criteria indicated that optimizing the response categories (Andrich, 1996; Linacre, 2002; Luo, 2005) would increase the validity and interpretability of the measures. Figure 1 indicates that, far from representing equal units of measurement, the rating categories vary markedly in the amounts of the variable they stand for. Categories 2 and 5 are never the most likely response, and category 4 dominates a much wider span of the central range of the measurement continuum than category 3. Though the observed scores tend to correspond with the modeled expectations, there is significant room for improvement.

Therefore, the original six-category response structure was reduced to a three-category response structure by combining all three categories of disagreement (very strongly disagree, strongly disagree, disagree) into one category, and by combining the two more extreme categories of agreement (strongly agree, very strongly agree) into a single category. The recoding resulted in a closer correspondence of observed and expected scores, as shown in Figure 2.

### Evaluation of Item Fit Statistics

Two items had markedly higher mean square fit statistics than the other items. These two items also plainly addressed off-construct topics not included in the scope of schools' responsibilities for facilitating parent involvement. For these reasons, the two items were removed from the scale. Table 3 shows the full text of the remaining 78 items.

Overall, the point-measure correlations were all positive and ranged from .54 to .76. With one exception, both the information-weighted and outlier-sensitive item mean square (infit and outfit, respectively) statistics ranged from 0.70 to 1.66, indicating data consistency of the quality required for meaningful interpretations of rating scale measures (Linacre, 2003b; Smith, 2002; Smith, 1996b, 2000).

### Measure and Calibration Reliability

Reliability is traditionally defined as the proportion of variance that truly expresses the genuine values measured, relative to the proportion due to independent and random error (e.g., Guilford, 1965, pp. 439-40). This definition is based on a statistical model of inter-variable correlations that does not distinguish the internal consistency of data from measurement error (Green, Lissitz, & Mulaik, 1977; Sijtsma, 2009). A measurement-theoretical basis for reliability as within-variable sufficiency is expressed in terms of the separation statistic,  $G$ , which indicates the genuine spread of items or persons in standard error units distinct from data consistency indexes (Bond & Fox, 2007; Wright & Masters, 1982, pp. 92, 106).

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, and the ratio of that adjusted standard deviation to the root mean square error is the separation statistic,  $G$ . In the context of  $G$ , more variance relative to error translates directly into a higher estimate of instrument reliability. The separation index is typically more conservative than Cronbach's alpha, as  $G$  accounts for higher error associated with off-target instruments and data inconsistencies, information not available to alpha. A summary of the scaling results for the six- and three-category versions of the SEPPS is provided in Table 4. This table also shows the summary statistics for SEPPS configurations with 25 items per case, and for the 25 items ultimately recommended for general use. Results for these two latter versions are shown because, with about one-third of the data missing, the reliability coefficients obtained with an average of 25 items per case provide reasonable grounds for the expectation that similar reliabilities will be obtained for the recommended 25-item version.

The separation reliabilities of .95 (modeled) and .94 (real, or fit-deflated) for the 2,634 SEPPS person measures relative to the 78 items scaled are associated with ratios of true variation to error of about 4.00 to 4.25. These statistics indicate that respondents can be separated into at least seven statistically distinct levels along the measurement continuum (Wright & Masters, 1982, pp. 92, 105-6, 111-3). Cronbach's alpha (KR20) for the parent raw scores was .95, with a confidence interval (Fan & Thompson, 2001) of less than .01.

Calibration separation reliability for the items relative to the sample of parents surveyed was 1.00, with separation statistics  $G$  of 16.4 and 17.5, indicating that the items can be separated into about 22 statistically distinct ranges along the measurement continuum, with 94% of the variance being true variance.

Using Linacre's (1993) reliability nomograph as a guide, item subsets were analyzed so as to ascertain the minimum scale length yielding measurement reliabilities of at least .90. As long as measures were based on an average of 25 items per respondent and represented the full range of item calibrations, reliability coefficients were consistently greater than .90.

### Evaluation of Scale Invariance

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, form (A vs. B), language (English vs. Spanish), and mode of administration (respondent read the survey independently vs. survey was read to the respondent). Overall, correlations were all greater than .80, with the vast majority over .90, and many approaching 1.00 when disattenuated (Muchinsky, 1996). Disattenuation does not improve the quality of the measures, and is no substitute for improved measurement precision when such improvement is called for. Disattenuation does, however, enable a distinction between correlations that are low because they are affected by measurement error, versus those that are low because there is less of an association between the two sets of measures.

The six correlations between separately estimated item calibrations for four different subsamples based on child ages ranged from .92 to .99. The correlation of the calibrations produced by data from parents who filled out the form independently with those produced by parents who had the survey read to them was .98.

All 78 items were included in each comparison. Figure 3 is a scatter plot of the calibrations produced from separate analyses of responses from the parents of children age 5 versus the parents of children ages 11-13. This is a typical plot, with a correlation of .92 (disattenuated, 1.00). Figure 4 is a scatter plot of the calibrations produced from separate analyses of responses from parents who used the English-language forms and those who used the Spanish-language forms. The correlation for these calibrations,  $r = .83$  (.93, disattenuated), was the lowest found in these data, suggesting the presence of potentially significant cultural differences.

Mirroring these analyses of item calibrations based on different subsamples of parents, measures for the responding parents were estimated using different subsets of items. Figure 5 plots the measures estimated using items that appeared only on one form of the survey against measures estimated using items that appeared on both forms. That is, two entirely different sets of items

were used to measure one set of parents, and the pairs of measures were plotted. The linear plots illustrate that a constant unit size is obtained irrespective of the particular items used. No respondents were omitted from the scaling analyses or the comparisons. The correlation for the measures from the single-form items versus the measures from the both-forms items is .97 (1.00, disattenuated).

Both a Principal Components Analysis (PCA) of the raw scores in SPSS and a PCA of the standardized residuals in Winsteps indicated that about 55% of the variance was explained by the measures. The eigenvalue of the second variance component was 1.55, with 7.4% of the variance; no other components had eigenvalues greater than 0.88 or variances explained over 4.2%. In the Winsteps PCA, the remaining unexplained variance was apportioned among five contrasts, the largest of which accounted for only 3.1% of the total variance. The five contrasts in total amount to 10.2% of the total variance, and 22.6% of the unexplained variance.

The first residual factor separated the items describing attributes and actions of schools, e.g., "The school gives me choices with regard to services that address my child's needs," from those describing attributes and actions of teachers and administrators, e.g., "Teachers treat me as a team member," "Teachers and administrators at my child's school value my ideas and input." The next four factors, all very small, were not clearly interpretable. Nine of the 78 items had loadings greater than .40 in the first contrast, with four in the second, four in the third, one in the fourth contrast, and none in the fifth. Six of the loadings were over .50; the highest was .57.

To test the possibility that the items with positive and negative factor loadings in the first contrast measured separate constructs, the 22 items with positive loadings greater than .20 were used to generate unconstrained measures on all available cases (2,616), as were the 18 items with negative loadings less than -.20 (2,634 cases). Given the lower measurement reliabilities (.83 and .89, respectively) obtained with these shorter scales, a lower correlation is expected and obtained (.71), and disattenuation has a larger effect (.98).

Overall, these results reproduce the other tests of invariance and indicate that the scale is essentially unidimensional (Linacre, 2003a; Smith, 1996a, 2002).

### Standard Setting

Given the successful calibration of the instrument, a basis for setting an objective standard is established. The standard setting task has three major components (Grosse & Wright, 1986; Stone, 2001; Stone, Beltyukova, & Fox, 2008). Standard setting in the usual context of educational testing requires 1) a decision as to how far up the scale the standard has to be for it to be an acceptable representation of achievement; 2) a decision as to what should be the minimally acceptable likelihood of success at the level decided in step 1; and 3) a decision as to whether the lower or upper bound of the error range

around the 95% confidence interval determines the minimum quantitative standard at which the cut-off will be set.

This procedure had to be modified slightly in the context of the parent survey. Educational achievement standards are based on examination items typically scored dichotomously, as correct or incorrect, and the standard setting procedure is then focused on adding a given mastery level (60% or 75%, for instance) and confidence level to the criterion level. Mastery in the context of the parent surveys is better conceived in terms of the rating scale response probabilities than in terms of a minimum proportion of correct responses.

Stakeholders unanimously agreed that schools would not effectively meet the terms of the SPP indicator 8 unless parents expressed agreement with all items up to and including the item that calibrated at 600 on the scale (item 131, "The school explains what options parents have if they disagree with a decision of the school;" step 1). Stakeholders further agreed that the minimal acceptable likelihood of agreement with the item calibrating at 600 should be 0.95 (step 2). Finally, it was agreed that the cut-score, instead of being set at one of the bounds of the confidence interval, should be set equal to the calibration of the threshold item (step 3), in order to make reporting on the indicator (which requires calculation of the percent at or above the standard) as straightforward as possible for state entities. The SPP indicator 8 percentage for a given state is then determined by dividing the number of parents with measures at or above 600 by the total number of parents with valid measures. Additional details on the standard setting process and on the setting of SPP indicator 8 targets are available (Fisher, Elbaum, & Coulter, 2010).

## Discussion

The emergence of a consistent ordering of items across multiple independent samples of respondents provides compelling evidence for the existence of a single construct underlying the items. Beyond having appropriate item content, a measurement tool should also yield invariant measures, that is, measures that mean the same thing (within the range of error) for all individuals within a population, and for all samples within the population to which the tool will be applied. If most item response patterns differed across subgroups of respondents, for example, parents of younger children versus parents of older children, or parents of Hispanic origin versus parents of African-American origin, this variability would invalidate scale comparisons across those groups. The demonstration of invariance – not in the absolute level of the trait being measured, but in the pattern of responses – for different groups of respondents provides the strongest support for the validity of the measures.

The invariant consistency of the item order defines a hierarchy of interrelated and increasingly challenging (for schools) ways of partnering with parents. This hierarchy effectively defines a path for the development of schools' partnership efforts. For instance, schools providing information in a language the

parent understands (calibrating at 499) and teachers being available to speak with parents (511) are likely necessary conditions for parents to feel treated as a team member (528) or to feel part of the decision-making process (541). The item hierarchy in Table 3 could be interpreted as a map (Bond & Fox, 2007; Wilson, 2005) indicating the direction in which efforts should be guided to facilitate improved school-parent relationships.

The existence of a calibrated bank of items shown to measure schools' efforts to partner with parents of children served under IDEA creates new opportunities for rigorous assessment and for comparable accountability standards for states. This research enables states to select for administration subsets of items deemed especially appropriate to their needs, without compromising the comparability of their measures against any other states' measures. In principle, the items selected need not be restricted to those in the calibrated bank as long as construct and content validity are demonstrated in pilot testing of the items (Bond & Fox, 2007; Wilson, 2005).

Further, states have the option of employing the entire SEPPS item bank in adaptive administrations, custom tailoring the survey to local schools' or individual parents' needs. Of course, if this is done, care must be taken to ensure that ongoing standards for reliability and validity are met.

## Limitations

We note three limitations of this study. First, the invariance of the SEPPS is not perfect; further analyses investigating differential test, item, and person functioning (Linacre, 2003b; Karabatsos, 2003; Smith, 1996a, 1996b, 2000) may yield interesting insights across specifically defined populations. As an example, the lowest correlation between separate-sample calibrations ( $r = .83$ ) was obtained in the comparison of responses from parents who used the English-language forms and those who used the Spanish-language forms. Of the 109 persons who filled out the Spanish-language forms, 102 self-identified as Hispanic or Latino, and four as American Indian or Alaskan Native. About 72% of all parents reporting Hispanic or Latino ethnicity filled out an English-language survey. About two-thirds of the parents filling out the Spanish-language version of the form were from New Mexico, and about 30% were from Georgia.

There were no statistically or substantively significant differences in the average measures, errors, or fit statistics for the English-language vs. Spanish-language respondents. However, the plot of the calibrations obtained from the two groups (see Figure 4) reveals some interesting differences in their response patterns. For example, parents responding to the Spanish-language forms were significantly more likely than parents responding to the English-language forms to agree with the following items: "I was offered the services of an interpreter, if needed," "I was offered special assistance (such as child care or transportation) so that I could participate in the IEP meeting," "We discussed how my child would participate in statewide assessments," and "In preparation for my child's

transition planning meeting I was given information about options my child will have after high school.”

Conversely, parents responding to the Spanish-language forms were significantly less likely than parents responding to the English-language forms to agree with the following items: “My child’s school offers parents a variety of ways to communicate with teachers (face-to-face meetings, email, phone, etc.),” “I (would) feel comfortable expressing disagreement with school staff,” “Information is provided to me in a language I understand,” and “Written information I receive is understandable.” When the eight items that function differentially across language forms are omitted from the comparison, the remaining 70 item calibrations correlate .91, indicating the basis for comparable interpretations of measures estimated from these items.

A second limitation of the study has to do with the applicability of the scale to parents of children receiving preschool special education services. Although the pattern of responses of parents of preschool children was in every way similar to that of parents of older children (cf. Figure 3), additional feedback solicited by NCSEAM from stakeholders involved in preschool special education, as a follow-up to the initial validation study, indicated the advisability of modifying the items slightly for parents of preschool children. In particular, it was pointed out that not all children receiving preschool special education services receive those services in school settings, hence references to schools, administrators, etc. would not have universal applicability to all parents in this sub-population. Stakeholders further recommended that the version of the scale administered to parents of preschool children should include at least some items of particular relevance to parents of children in this age group, for example, items related to families’ transition from early intervention services (provided to families of children birth to three under Part C of the IDEA) to preschool special education services (provided to children ages 3-21 under Part B of the IDEA). Responding to these recommendations, NCSEAM developed and piloted additional items which scale together with the original items and are available for states’ use for this population (see <http://www.accountabilitydata.org/FamilyInvolvement619Survey.htm>).

A third, more minor, limitation of the study has to do the decision to use the child’s age, instead of grade in school, as a criterion for creating the dataset used for the analyses reported in this paper. As noted earlier, preschool children are generally between the ages of three and five; however, some 5- and 6-year olds, particularly children with severe developmental delays, may still be attending preschool. Since the data on children’s ages and grade levels were obtained through parents’ self-report, we felt that parents’ report of their child’s age was likely to be more accurate than their report of the child’s grade level. However, in using age as the selection variable, we may have included a small number of parents of children receiving preschool services in the sample that provided data for this paper. At the same time, for reasons discussed in the previous paragraph, there is evidence that the inclusion of a small number of

parents of children still in preschool does not appear to have affected the results.

## Conclusion

The findings of this study support the conclusion that the SEPPS yields valid and reliable measures the amount of effort that schools expend to facilitate meaningful involvement of parents of children receiving special education services. Statewide distributions of parents' measures on the SEPPS, based on a representative sample of parents of students receiving special education services, provide a sound foundation for application of a state-selected standard and reporting of the resulting percent of measures that are at or above that standard.

According to an analysis of states' Annual Performance Reports, conducted under the auspices of OSEP (Regional Resource Center Program, 2008b), some version of the SEPPS—most typically, the 25-item version recommended by NCSEAM and available at [www.accountabilitydata.org](http://www.accountabilitydata.org)—was being used by 35 states and territories (representing 58% of the U.S. Department of Education's 60 reporting entities) to address Indicator B-8. Of these states, 18 had also adopted the NCSEAM-recommended standard for accountability reporting.

Use of a rigorous measurement tool to evaluate performance on the SPP indicators affords states an additional advantage that is of considerable practical value. The fact that efforts to facilitate parent involvement can be measured, using the SEPPS, in a common metric across different districts means that states can reliably differentiate between local programs that are performing well and those that are performing less well on this indicator. States are in fact required to report annually to the public on the performance of each school district on the SPP indicators. By identifying districts whose efforts to involve parents are worthy of emulation, as well as those in need of improvement, states can use their SPP data to target their technical assistance where it is most needed.

## References

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- 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.
- Bond, T., & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*, 2nd edition. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Data Accountability Center. (2010a). Table 1-3. Students ages 6 through 21 served under IDEA, Part B, by disability category and state: 2004. Retrieved from <https://www.ideadata.org/PartBData.asp>
- Data Accountability Center. (2010b). Table 1-16. Students ages 6 through 21 served under IDEA, Part B, by race/ethnicity and state: 2004. Retrieved from <https://www.ideadata.org/PartBData.asp>
- Fan, X., & Thompson, B. (2001). Confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement*, 61(4), 517-531.
- Fisher, W. P., Jr., Elbaum, B., & Coulter, A. (in preparation). Stakeholder input in scale development, item validation, and standard setting: An example from special education accountability.
- Fisher, W. P., Jr., & Wright, B. D. (Eds.). (1994). Applications of probabilistic conjoint measurement. *International Journal of Educational Research*, 21(6), 557-664.
- Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37(4), 827-833.
- Grosse, M. E., & Wright, B. D. (1986). Setting, evaluating, and maintaining certification standards with the Rasch model. *Evaluation & the Health Professions*, 9(3), 267-285.
- Guilford, J. P. (1965). *Fundamental statistics in psychology and education* (4th Ed). New York: McGraw-Hill.
- Individuals with Disabilities Education Act of 2004, P. L. No. 108-446, 118 Stat. 2647.
- Karabatsos, G. (2003). A comparison of 36 person-fit statistics of Item Response Theory. *Applied Measurement in Education*, 16, 277-298.
- Linacre, J. M. (1993). Rasch-based generalizability theory. *Rasch Measurement Transactions*, 7(1), 283-284; [<http://www.rasch.org/rmt/rmt71h.htm>].
- 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), 942-943.

- 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. (2009). *A user's guide to WINSTEPS Rasch-Model computer program*, v. 3.68.1. Chicago, Illinois: 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), 443-455.
- Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, 56(1), 63-75.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (Reprint, with Foreword and Afterword by B. D. Wright, Chicago: University of Chicago Press, 1980). Copenhagen, Denmark: Danmarks Paedagogiske Institut.
- Regional Resource Center Program. (2008). Indicator Support Grid. Retrieved from <http://spp-apr-calendar.rfcnetwork.org/getfile/view/id/139>
- Regional Resource Center Program. (2008). SPP/APR Summary Document (August, 2008). Retrieved from (<http://spp-apr-calendar.rfcnetwork.org/explorer/view/id/417/?4#category4>
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120.
- 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.
- Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement*, 1(2), 199-218.
- Stenner, A. J. (2001). The Lexile Framework: A common metric for matching readers and texts. *California School Library Journal*, 25(1), 41-42.
- Stone, G. (2001). Understanding Rasch measurement: Objective standard setting (or truth in advertising). *Journal of Applied Measurement*, 2(2), 187-201.
- Stone, G. E., Beltyukova, S. A., & Fox, C. M. (2008). Objective standard setting for judge-mediated examinations. *International Journal of Testing*, 8(2), 180-196.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago, Illinois: MESA Press.
- Wright, B. D., & Stone, M. H. (2004). *Making measures*. Chicago: Phaneron Press.

Table 1

*Target Number of Participants by Ethnicity and Age Group for the State of New Mexico*

---

Ethnicity	Age Group			
	5	6-10	11-13	14-21
White	15	50	50	50
African American	0	0	0	0
Hispanic or Latino	25	100	100	100
Native American	20	100	100	100
Asian or Pacific Islander	0	0	0	0
Total	60	250	250	250

---

Table 2

*Characteristics of Participants*

Characteristic	<i>n</i>	Percent of Sample	Percent of Children. Ages 6-21 Served, by Ethnicity and Category of Disability
<b>State of Residence</b>			
Florida	111	4.2	
Georgia	683	25.9	
Mississippi	130	4.9	
New Hampshire	793	30.1	
New Jersey	408	15.5	
New Mexico	507	19.2	
Unknown	2	0.1	
<b>Ethnicity</b>			
White	1469	55.8	58.7

Black or African American	504	19.1	20.5
Hispanic or Latino	417	15.8	17.2
Asian or Pacific Islander	60	2.3	2.1
American Indian or Alaskan Native	127	4.8	1.5
Multi-Racial	25	0.9	n/a
Unknown	32	1.2	
Child's Age Group			
5 (kindergarten)	158	6.0	
6-10 (elementary school)	956	36.3	
11-13 (middle school)	601	22.8	
14-21 (high school)	919	34.9	
Child's Category of Disability			
Autism	336	12.8	2.7
Deaf/Blindness	16	0.6	<0.01
Emotional Disturbance	138	5.2	7.9
Hearing Impairment	60	2.3	1.2

Mental Retardation	197	7.5	9.2
Multiple Disabilities	168	10.4	2.2
Orthopedic Impairment	52	2.0	1.1
Other Health Impaired	359	13.6	8.4
Specific Learning Disability	577	21.9	46.4
Speech or Language Impaired	498	18.9	18.8
Traumatic Brain Injury	61	2.3	0.4
Visual Impairment or Blindness	67	2.5	0.4
Developmental Delay	n/a <sup>a</sup>	n/a <sup>a</sup>	1.2
Survey Form Completed			
Form 1 (158 items)	1292	49.1	
Form 2 (160 items)	1342	50.9	
Mode of Administration			
Completed the survey independently	1831	69.5	
Completed the survey while facilitator read items aloud	660	25.1	
Unknown	143	5.4	

### Language in which the Survey was Completed

English	2161	82.0
Spanish	109	4.1
Another language	5	.2
Unknown	359	13.6

---

Note: Data on state of residence, ethnicity, child's age, child's category of disability, and mode of survey administration are based on parents' responses on the survey forms.

<sup>a</sup>The category of "Developmental Delay" is a relatively new category of eligibility for children ages birth through 9. The survey form used for the field study did not include this option.

Table 3

*SEPPS Items in Calibration Order*

Item calibration	Item text
702	My child's school provides funding, transportation, or other supports for parents to participate in training workshops.
687	My child's school invites parents to assist in training school staff.
687	The school or district holds public meetings to gather parent input on special education services.
670	My child's school connects families to other families that can provide information and mutual support.
664	The school offers parents training about special education issues.
661	My child's school offers parents training if they need help understanding the curriculum being taught to their child.
661	My child's school has helped me find resources in my community such as after-school programs, social services, etc.
659	My child's school connects parents to organizations that serve parents of children with disabilities.
657	My child's school provides parents with training on how to resolve disagreements using processes like conciliation and mediation.
652	The school gives parents the help they may need to play an active role in their child's education.
650	My child's school has support groups for parents.
647	I was given information about organizations that offer support for parents of students with disabilities.
645	My child's school encourages parents to organize support groups.

643 My child's school told me how to request therapies that my child needs.

641 I was offered special assistance (such as child care) so that I could participate in the Individualized Educational Program (IEP) meeting.

640 I have been asked for my opinion about how well special education services are meeting my child's needs.

640 In preparation for my child's transition planning meeting I was given information about options my child will have after high school.

638 The school provides information on agencies that can assist my child in the transition from school.

628 The school gives me choices with regard to services that address my child's needs.

626 I was given information about the research that supports the instructional methods used with my child.

600 The school explains what options parents have if they disagree with a decision of the school.

596 Written justification was given for rejecting any of my proposals.

591 I was given information about the continuum of alternative placements.

589 The school communicates regularly with me regarding my child's progress on IEP goals.

585 I was given information about the curriculum and materials used with my child.

583 I was given information about the instructional methods used with my child.

583 My child's school gives me enough information to know whether or not my child is making adequate progress.

581 I was given information about expectations about what my child will be able to achieve.

581 I was given all reports and evaluations related to my child prior to the IEP meeting.

576 At the IEP meeting, we discussed how my child would participate in statewide assessments.

576 The school has a person on staff who is available to answer parents' questions.

576 Teachers and administrators at my child's school invite me to share my knowledge and experience with school personnel.

573 Written justification was given for the extent that my child would not receive services in the regular classroom.

567 Teachers and administrators at my child's school consider what my hopes and dreams are for my child.

567 Teachers and administrators seek out parent input.

566 I feel I can disagree with my child's special education program or services without negative consequences for me or my child.

563 Teachers and administrators at my child's school answered any questions I had about Procedural Safeguards.

563 I was offered the services of an interpreter, if needed.

561 Teachers and administrators ensure that I have fully understood the Procedural Safeguards [the rules in the federal law that protect the rights of parents].

561 We discussed whether my child needed services beyond the regular school year.

559 I am considered an equal partner with teachers and other professionals in planning my child's program.

559 All of my concerns and recommendations were documented on the IEP.

558 I was given enough time to fully understand my child's IEP.

556 My child's teachers are good at working with families.

554 Teachers and administrators show sensitivity to the needs of students with disabilities and their families.

552 I was given information about my child's educational needs.

552 Teachers and administrators at my child's school expect parents to participate in decision-making.

552 Teachers and administrators at my child's school value my ideas and input.

551 The evaluation results were thoroughly explained to me.

549 The school principal is available when I need his/her help concerning my child.

554 We discussed whether my child could be educated satisfactorily in the regular classroom with appropriate aids and supports.

542 The school offers parents a variety of ways to communicate with teachers (face-to-face meetings, email, phone, etc.).

541 I felt part of the decision-making process.

541 Teachers and administrators at my child's school willingly accept information that I offer about my child.

540 Teachers and administrators encourage me to participate in the decision-making process.

537 I (would) feel comfortable expressing disagreement with school staff.

536 My child's evaluation report is written in terms I understand.

535 I have a good working relationship with my child's teacher's.

534 I was given information about the behavior and discipline policies of my child's school.

534 Teachers and administrators at my child's school acknowledge parents' efforts.

534 I feel welcome when I come into my child's classroom.

532 At the IEP meeting, we discussed accommodations and modifications that my child would need.

531 My child's teachers give me enough time and opportunities to discuss my child's needs and progress.

531 Teachers and administrators at my child's school respect my family's values.

531 IEP meetings are scheduled at a time and place that are convenient for me.

528 I was given information about my rights as addressed in the Procedural Safeguards.  
528 Teachers treat me as a team member.  
525 I am comfortable asking questions and expressing concerns to school staff.  
523 Teachers and administrators at my child's school maintain the confidentiality of students and families.  
521 Teachers and administrators respect my cultural heritage.  
519 Written information I receive is written in an understandable way.  
516 My child's teachers answer my questions.  
512 I was given adequate notification of upcoming IEP meetings.  
512 I was given information about my child's eligibility for and placement in special education.  
511 Teachers are available to speak with me.  
509 Everyone at the IEP meeting was introduced to me.  
502 I was given information about my rights as a parent of a child who is eligible for special education services.  
499 Information is provided to me in a language I understand.

---

Table 4

*Summary Statistics for the Original 6-Category SEPPS, the Optimized 3-Category SEPPS, a 40-Item SEPPS, and the Recommended 25-Item SEPPS*

Statistic	Original 6-Category SEPPS	Optimized 3-Category SEPPS	40-Item SEPPS (25 items per case)	Recommended 25-Item SEPPS (16 items per case)
<b>Sample sizes</b>				
Items	80	78	40	25
Non-extreme measures	2600	2468	2419	2383
All measures	2634	2634	2634	2633
<b>Mean/SD/Error</b>				
Non extreme measures	53.0/ 12.6 / 1.8	498 / 102 / 19	495 / 95 / 24	494 / 96 / 30
All measures	53.8 / 14.4 / 2.0	491 / 133 / 24	492 / 130 / 30	490 / 129 / 37
<b>Mean item count / score</b>				

Non-extreme measures	47.1 / 180.3	45.8 / 90.4	24.7 / 50.4	16.1 / 31.8
All measures	47.1 / 181.5	45.8 / 89.4	24.7 / 50.0	16.1 / 31.6
Measurement Separation G/Reliability (Cronbach/Rasch)				
Non-extreme measures, real	5.8 / .96 / .97	5.0 / .95 / .96	3.6 / .92 / .93	2.9 / .90 / .91
Non-extreme measures, modeled	6.7 / .98	5.5 / .97	3.9 / .94	3.2 / .89
All measures, real	4.8 / .96	4.1 / .94	3.3 / .92	2.8 / .89
All measures, modeled	5.1 / .96	4.2 / .95	3.4 / .92	3.0 / .90
Calibrations Mean/SD/error				
Mean/SD/error	50 / 5.2 / .3	576 / 52 / 3	559 / 49 / 3	576 / 47 / 3
Mean counts and scores				
Respondents	1550.8	1546.4	1623.2	1695.6
Item sum score	5975.2	3017.3	3290.6	3324.7
Calibration mean square fit – Mean/SD				
Infit	1.00 / .25	1.03 / .22	0.97 / .25	1.00 / .22
Outfit	1.02 / .31	1.06 / .27	1.00 / .27	1.00 / .24

Calibration separation G and reliability

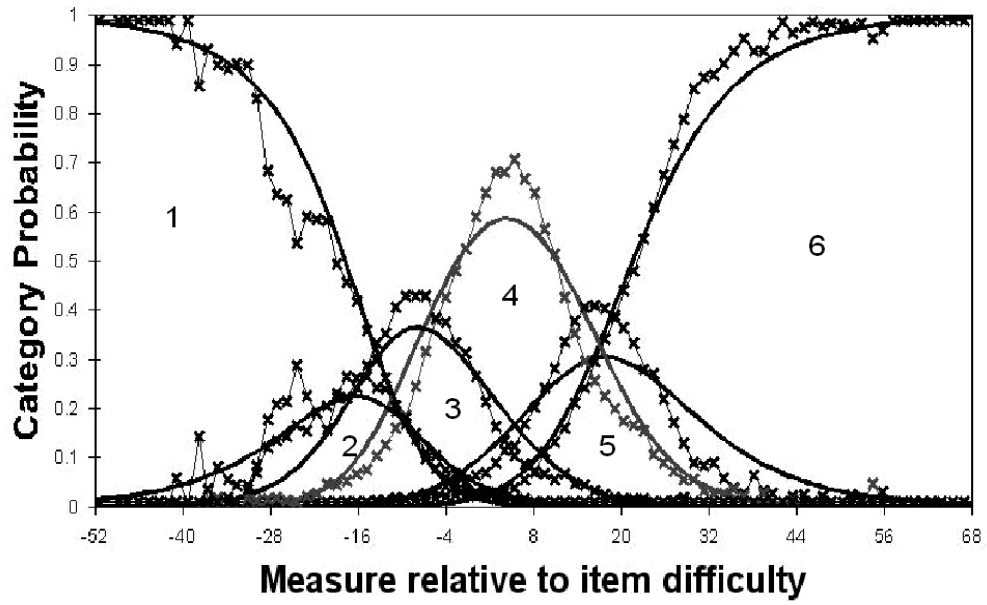
Real	16.5 / 1.00	16.4 / 1.00	15.9 / 1.00	15.1 / 1.00
Modeled	17.1 / 1.00	17.5 / 1.00	16.9 / 1.00	16.1 / 1.00

---

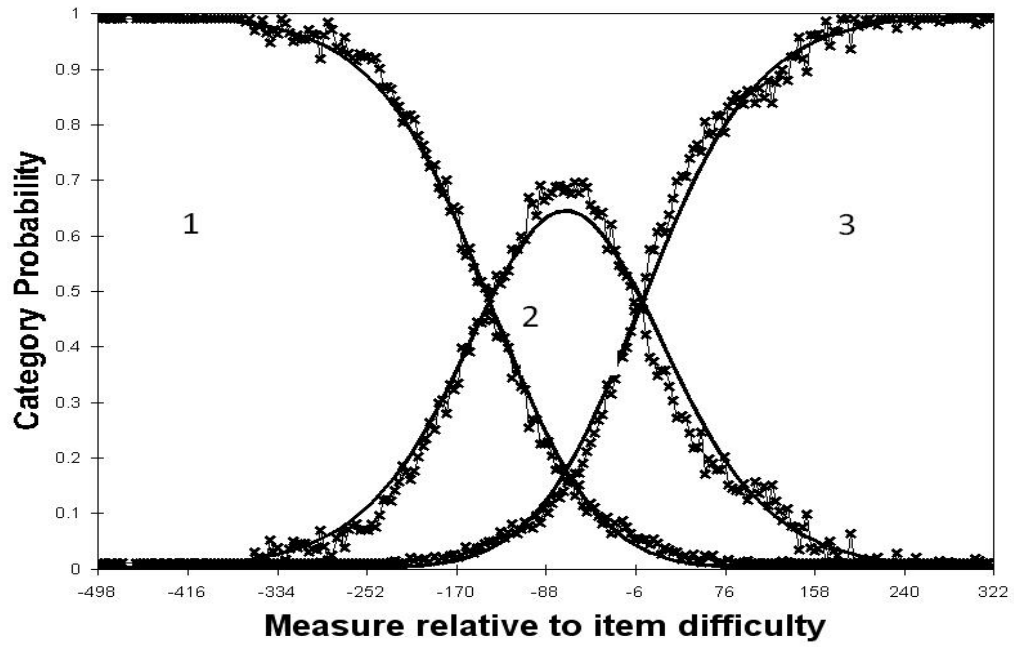
Note: The measures based on the three-category instrument are in the unit defined by the standard setting process, and are not comparable with measures using the original six categories. Cronbach's alpha is shown only once as it is estimated only once, on the basis of the raw scores.

“Non-extreme” indicates a summary of measures excluding extreme minimum and maximum scores.

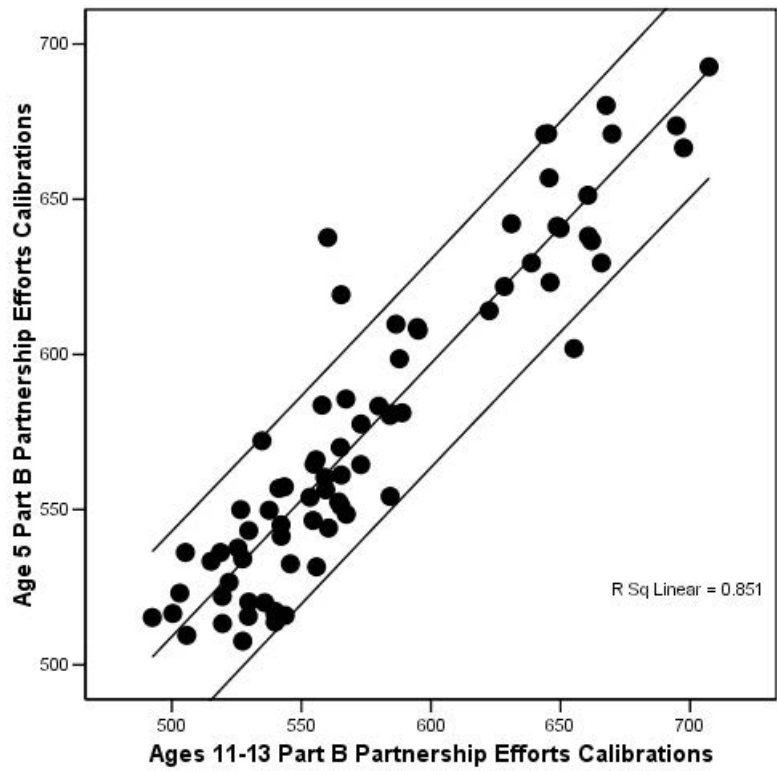
Real separation and reliability coefficients are inflated by model fit, whereas the modeled versions are not.



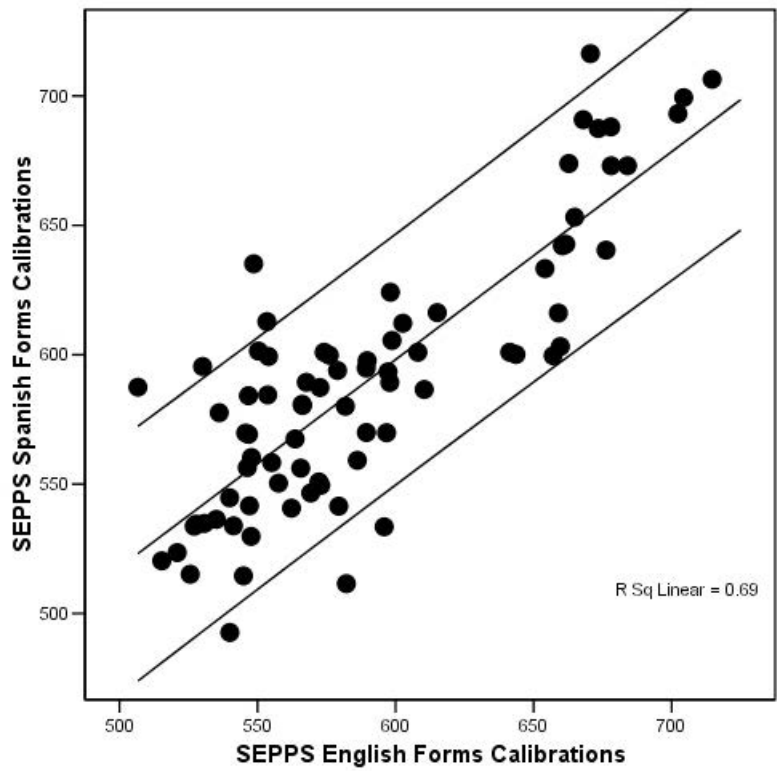
*Figure 1.* Original SEPPS response categories.



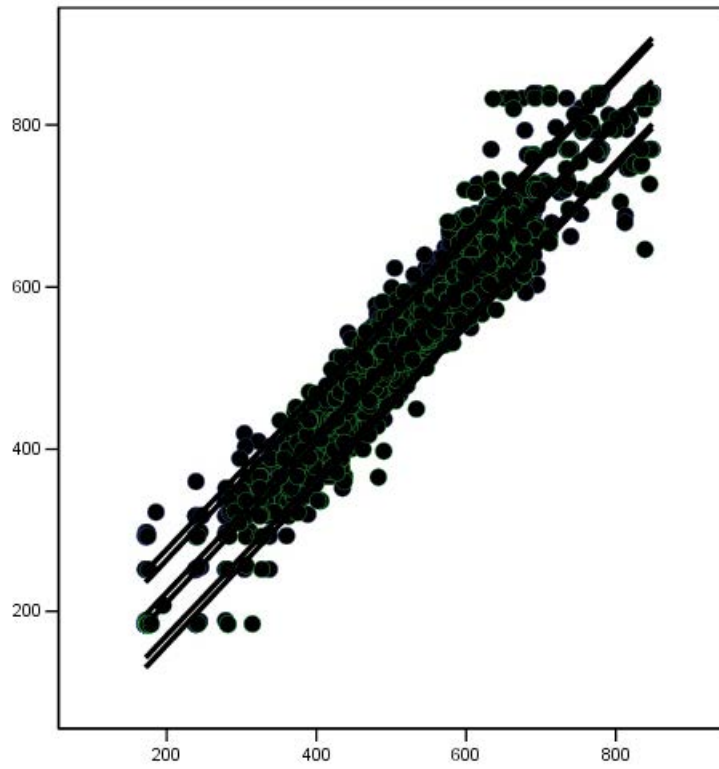
*Figure 2.* Optimized SEPPS response categories.



*Figure 3.* Calibrations from parents of children age 5 versus calibrations from parents of children ages 11-13 ( $r = .92$ ; 1.00, disattenuated).



*Figure 4.* Calibrations from parents using English-language form vs those who used the Spanish form ( $r = .83$ ; disattenuated, .93).



*Figure 5.* Measures from single-form items (horizontal axis) vs. measures from both-form items (for both comparisons,  $r = .97$ ; 1.00, disattenuated).