2005

Analyzing CSR Implementation with the Rasch Model

Susan Gracia
Rhode Island College, sgracia@ric.edu

Analyzing CSR Implementation with the Rasch Model

Susan Gracia
The Education Alliance at Brown University

Paper presented at the at the American Educational Research Association 2005 Annual Meeting, Montreal, Canada.

Background

> Comprehensive school reform is not just another school improvement strategy–it is a significant leap forward in reforming today's public schools. Comprehensive school reform addresses all students, ll academic subjects and all teachers. When done well, a school is overhauled from top to bottom. Adding one program on top of another is thrown out in favor of the much more difficult work of reorganizing schools, targeting professional development for teachers and principals, changing curriculum and making tough budget decisions (Education Commission of the States, 1999).

The objective of the federal Comprehensive School Reform (CSR) Program is to improve student achievement by supporting the implementation of comprehensive school reforms based on scientifically based research and effective practices so that all children, especially those in low-performing, high-poverty schools, can meet challenging state content and academic achievement standards (U.S. Department of Education, 2002). Specifically, comprehensive school reform is a means to improve student achievement through reorganizing and revitalizing the entire school, rather than implementing isolated programs. The CSR program is incorporated into The No Child Left Behind Act of 2001, with an emphasis on selection of appropriate strategies and practices that are based on scientifically based research and that address each of the following 11 components:

1. Effective, research-based, replicable methods and strategies
2. Comprehensive design with aligned components
3. Professional development
4. Measurable goals and benchmarks
5. Support within the school
6. Support for principals and staff
7. Parental and community involvement
8. External technical support and assistance
9. Evaluation strategies
10. Coordination of resources
11. Scientifically-based evidence of improved student achievement

Federal law also stipulates that the implementation and impact of CSR programs be evaluated. One means of evaluating the implementation of CSR is through the use of teacher surveys requiring teachers to rate their perception of their own and their school's implementation of the various components of CSR. As such, the development of high quality, valid scales for the measurement of CSR implementation is critical.

Scale Validation

The adequacy of teacher survey scales is typically assessed using classical test score theory analysis, including examinations of descriptive statistics, item correlations, factor analyses, and reliability analyses, among other procedures. However, the application of classical test theory to the evaluation of the measurement properties of scales has several drawbacks. Among these drawbacks are the following:

- Estimates of item difficulty and person ability are sample dependent
- Estimates of item difficulty cannot be compared unless the estimates come from the same sample or assumptions are made as to the comparability of different samples
- All items are assumed to be of equal importance in defining the construct
- Although ordinal, Likert scale ratings are often assumed to lie on an equal interval scale
- Respondent scores are summed ratings; consequently, people who assign very different ratings to the same items can be assigned the same total score, making it appear as if their
- Missing data is problematic (Smith, 2000; Bond and Fox, 2001)

The Rasch model, on the other hand, addresses many of the weaknesses of classical test score theory by resting on the idea that people and data conform to a hierarchy of "less than/more than" on a continuum, or variable, of interest. For example, some teachers have implemented CSR more than others, and some items describe higher or lower levels of CSR implementation than others. The model is also based on the assumption that some people are more likely to agree with easy items (e.g., depicting low level CSR implementation) than difficult to agree with items and that all items are more likely to be agreed with by high CSR implementers than low CSR implementers (Bond and Fox, 2001). The Rasch model uses the traditional total score (the sum of all item ratings) to specify "the occurrence of an event as a probability rather than certainty and [maintain] order such that the probability of providing a certain response defines an order or respondents and items" (Sampson and Bradley, 2003, p.6). The Rasch model exhibits several advantages over classical test score theory in that it parameters are neither sample nor test dependent, missing data are not problematic, total scores have interpretable meaning (i.e., they describe a respondent's place on the continuum of the variable), and ordinal data from Likert ratings to equal interval data, enabling researchers to interpret the size of the differences between people in their level of CSR implementation (Smith, 2000; Bond and Fox, 2001).

Research Objectives

The purpose of this research was to examine the measurement properties of a CSR Implementation Scale developed using classical test theory. Rasch analyses were employed to determine (1) the degree to which the scale meets the assumptions of the Rasch model; 2) the validity and reliability of the scale; 3) how respondents utilize the rating scale; 4) the nature of the continuum of CSR implementation; and 5) ways to optimize scale length, both in terms of eliminating redundancy and adding items where gaps in the continuum of the CSR implementation variable might occur.

Methods

Study Sample

Data were obtained from an evaluation study of CSR implementation in a large urban school district. Subjects were K-12 teachers in schools implementing CSR. There were 1645-2100 subjects (number of cases depended on analysis) who had usable data on the CSR Implementation Scale.

The vast majority of teacher survey respondents were elementary school teachers (62%), as opposed to middle school or high school teachers (21% and 17%, respectively). Sixty-nine percent of respondents were regular education teachers; in contrast, 12% taught special education, 5% were bilingual education or ESL teachers, and 5% identified themselves as Title 1 teachers. Sixty-four percent of survey respondents reported that they held full or permanent certification. Likewise, 29% of respondents held provisional certification. Fewer than 2% of teacher survey respondents reported that they were not certified at all. Similarly, 76% of survey respondents indicated that the content areas they taught matched the content areas in which they were certified all of the time. A closer look at these data revealed another interesting fact concerning the certification of teacher respondents. While 88% of elementary school teachers and 75% of middle school teachers indicated that they were certified for elementary and middle school, only 36% of teacher survey respondents in high schools reported being certified to teach high school.

While 8% of teacher survey respondents consisted of first year teachers, the number of years taught by teacher survey respondents ranged from one to forty-three years. The average number of years taught by survey respondents was 11.59 years, with 36% of teachers had six or fewer total years of teaching experience. The average number of years that teachers had taught in their current CSR school was just over 7 years. Likewise, 15% of survey respondents reported that they had taught in their current school for less than one year. Finally, three-quarters of teachers who responded to the survey were female. In addition, 51% of teacher respondents were white, 17% were African American, and 12% were of Hispanic origin.

Survey data also revealed that relatively few teacher survey respondents held school responsibilities outside of teaching. Similarly, only a small number of teacher survey respondents held school or reform-related leadership positions. More than 10% of respondents reported serving on teams related to school reform; however, 8% or fewer coordinated, facilitated, or headed up reform activities.

Scale Description

The CSR Implementation Scale consisted of 53 items designed to measure teachers' and schools' current level of CSR implementation. Scale items were designed to tap into all of the components of CSR, with the exception of Strategies to improve academic improvement. The scale also contained items reflecting the level of teacher collaboration and communication, a research-based indicator of school-level implementation of innovations (Elmore, 1996: Little, 1982; Berman & McLaughlin, 1977,1978).

The CSR Implementation Scale was structured primarily as a set of closed-response questions aligned with the above components. The bulk of the items came from a pre-tested, statistically validated instrument used by evaluators in other CSR-related research and evaluation projects. The scale combined field-tested indicators of CSR-related teacher practice, as well as research-based school- and district-level factors influencing the implementation of educational innovations. The goal of the survey was to assess school-level CSR implementation through the eyes of teachers.

Rasch Analysis

Rasch analyses were conducted using WINSTEPS software (Linacre, 2005). WINSTEPS begins with a central estimate for each person measure, item calibration, and rating scale category structure calibration. An iterative version of the PROX (normal approximation) algorithm is used reach a rough convergence to the observed data pattern. The JMLE (UCON) method is then iterated to obtain more exact estimates, standard errors and fit statistics. This implementation of the JMLE (UCON) (unconditional maximum likelihood, joint maximum likelihood) method uses proportional curve fitting for finding improved estimates. Measures are reported in Logits (log-odds units). Fit statistics are reported as mean-square residuals, which have approximate chi-square distributions and t standardized (Linacre, 1991-2005, p. 12).

Results

Model Fit

Fifty-three items made up the section of the CSR evaluation survey aimed at assessing CSR implementation. An initial Rasch analysis was run on all 53 items, and item fit statistics were examined to form the rationale for a possible decision to include just some of these items in subsequent analyses. The goal of this initial analysis was to avoid including items that showed evidence of inadequate fit and were not drawing on the same construct as the other items. The initial pool of 53 items is displayed in Figure 1.

1a. Our school reform efforts help our students meet district and state standards.
1b. Our school reform efforts help teachers to align curriculum and assessment with state standards.
1c. Our school reform efforts help teachers to meet the needs of all students.
1d. Our school reform efforts involve the entire school, including teachers and students from all grade levels and subject areas.
1e. Our school reform efforts help to align the different programs and strategies in our school.*
1f. Our school reform efforts are focused primarily on improving teaching and learning in our school.
1g. Our school reform efforts are based on the latest research and evidence on best practice.
2. Overall, teachers support the school's comprehensive school reform efforts.
3. Parents play an active role in the school's reform efforts.*
4. Parents understand and support the school's reform efforts.
5. District reform initiatives fit together effectively with the model we have adopted.
6. District administrators support building-level initiatives.
7. Our school's leadership team takes an active role in promoting school reform efforts.
8. The people in reform-related leadership roles work well together.
9. I support the comprehensive school reform approach that my school is using.
10. I am committed to maintaining the changes that I have made in my teaching.
11. When we adopted our current school reform strategy, it was necessary to change how we worked with students.*
12. Community members play an active role in our reform efforts.*
13. Community members understand and support our reform efforts.
14. The practices and strategies we are implementing form a coherent whole.*
15. Curriculum, instruction, and learning materials are well coordinated from one level to the next.
16. There is consistency in curriculum, instruction, and learning materials among teachers in the same grade.
17. The school schedule provides sufficient planning time to implement school reform practices and strategies.*
18. I look forward to collaborating with my colleagues. *
19. It is safe to voice our candid opinions on controversial or delicate topics. *
20. My colleagues and I openly and constructively discuss such topics. *
21. My colleagues are team players -- receptive to modifying their teaching so that all staff will be on the same page.
22. As a teaching staff we regularly assess our progress in implementing the model.
23. I understand our school goals.
24. I understand how our reform efforts are tied to school goals.
25. The administrators and teachers share responsibility for reform efforts.
26. Our school analyzes data and student assessments regularly to make appropriate curricular or instructional changes.
27. I communicate regularly with my students' parents. *
28. The strategies we use for communicating with families are effective. *
29. Our school provides information to families on how to monitor and discuss schoolwork at home.
30. Our school has strategies to communicate with parents who do not speak English well.*
31. I have received the support I need to implement our school's reform efforts.
32. Teachers in our school have an opportunity to talk with each other about school reform efforts.
33. Teachers in our school are able to express their concerns to the administration.
34a. Professional development opportunities have helped me understand how the reform strategies and practices help our students meet state standards. *
34b. Professional development opportunities have helped me understand the concepts and principles supporting the activities I am expected to implement. *
34c. Professional development opportunities have helped me understand the specifics of what is expected of me on a day-to-day basis. *
34d. Professional development opportunities have helped me understand how the new practices fit with each other. *
35a. Workshops and professional development opportunities have provided the skills and training needed to implement reform efforts. *
35b. Workshops and professional development opportunities relate to, build on and extend what we learned in previous sessions. *
35c. Workshops and professional development opportunities are directly tied to school improvement goals. *
35d. and professional development opportunities are attended by most, if not all, faculty members in the school. *
35e. Workshops and professional development opportunities have given me ideas and procedures applicable to many lessons.
35f. Workshops and professional development opportunities have increased my understanding of content concepts and principles.
35g. Workshops and professional development opportunities have helped me learn how to use research-based practices.
35i. Workshops and professional development opportunities have increased the consistency of instructional strategies used by teachers. *

* Item deleted from subsequent analyses.

Figure 1: Initial Item Pool

Using guidelines regarding reasonable item mean square ranges for infit (Bond and Fox, 2001), underfitting items with mean square infit statistics greater than 1.3 (representing too much, unpredictable variation) and overfitting items with mean square infit statistics less than 0.75 (indicating too little variation) were highlighted. Twenty-one misfitting

items were identified and eliminated, and subsequent analyses were performed with only the 32 fitting items.[1]

Validity of a scale can be measured by the degree to which the Rasch model can predict the response of each teacher to each item. Fit statistics are then used to test the validity of the estimated measures. In the Rasch model, the difference between the response expected by the model and the observed response is a residual. Fit statistics are normalized mean squared residuals (across items for each person or across persons for each item). These mean-square fit statistics are reported in two different ways: infit and outfit. The infit statistic (information-weighted fit statistic) emphasizes residuals for items that are close to the person's ability. The outfit statistic (outlier-sensitive fit statistic) reflects large differences between observed and expected values for items that are far from the person's ability.

With an expected value of 1, infit and outfit mean squares show the amount of the distortion of the measurement system within the scale. Values smaller than 1.0 indicate observations that are too predictable and overfit the Rasch model. Values greater than 1.0 indicate unpredictability, umodeled noise, or model underfit. Zstd statistics are t-tests of they hypothesis, "Do the data fit the model (perfectly)?" As t-statistics, their expected value and standard deviation are 0 and 1, respectively. The Zstd statistics are considered to be consistent with the model if their distribution agrees with the expectations of the standard normal distribution (within plus or minus 2 standard deviations of the mean). If mean square statistics are acceptable, the Zstd statistic can be ignored because mean squares near 1 indicate little distortion of the measurement system, regardless of Zstd value (Linacre, 1991-2005, p. 243).

Person and item fit statistics are generated in any Rasch analysis. Person fit describes how consistent a person's response pattern is with how s/he is expected to respond. When the items a person answers have been calibrated along a variable from easy to hard, that person's response pattern is expected to be consistent with the difficulty order of those items. Likewise, item fit statistics identify the degree to which responses to any item fit the model.

Bond and Fox (2001) provide guidelines for the interpretation of mean square and Zstd fit statistics. Under these guidelines, mean square statistics with values between .75 and 1.3 are considered to fit the Rasch model. As stated above, Zstd statistics greater than -2.00 and less than 2.00 also fit the Rasch model. An examination of the data in Table 1 below, reveals that the 32 item CSR Implementation Scale fits the Rasch model. The mean square infit and outfit statistics for persons are 1.02 and 1.00, well within the .75-1.3 range recommended by Bond and Fox. Similarly, the mean square infit and outfit

[1] Eight of the 21 deleted items focused specifically on professional development that teachers had received. Three others emphasized changes teachers had observed in their students. The remaining deleted items addressed a range of CSR implementation topics.

statistics for items are 1.01 and 1.00. In each case, the expected mean square value of 1.00 is achieved.[2]

Table 1: Person and Item Fit

SUMMARY OF 2094 MEASURED (NON-EXTREME) PERSONS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	77.9	31.4	.85	.27	1.02	-.3	1.00	-.4
S.D.	25.6	2.1	1.60	.08	.63	2.4	.69	2.4
MAX.	126.0	32.0	6.09	1.02	5.49	9.9	9.57	9.9
MIN.	3.0	3.0	-4.41	.22	.07	-6.9	.07	-6.9

REAL RMSE	.31	ADJ.SD	1.57	SEPARATION 5.12	PERSON RELIABILITY .96
MODEL RMSE	.28	ADJ.SD	1.58	SEPARATION 5.67	PERSON RELIABILITY .97

S.E. OF PERSON MEAN = .04

MINIMUM EXTREME SCORE: 5 PERSONS
LACKING RESPONSES: 1 PERSONS
VALID RESPONSES: 98.1%

SUMMARY OF 32 MEASURED (NON-EXTREME) ITEMS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	5097.9	2055.1	.00	.03	1.01	-.2	1.00	-.6
S.D.	724.0	14.6	.65	.00	.19	5.3	.23	5.3
MAX.	6259.0	2082.0	1.74	.03	1.47	9.9	1.69	9.9
MIN.	3131.0	2020.0	-1.14	.03	.83	-5.8	.76	-6.7

REAL RMSE	.03	ADJ.SD	.65	SEPARATION 20.11	ITEM RELIABILITY 1.00
MODEL RMSE	.03	ADJ.SD	.65	SEPARATION 20.90	ITEM RELIABILITY 1.00

S.E. OF ITEM MEAN = .12

UMEAN=.000 USCALE=1.000

Reliability

In the Rasch model, reliability, for both persons and items, refers to the percentage of observed responses that are reproducible. The person measure reliability estimates how well people are discriminated based on their estimated CSR implementation. According to Linacre (1991-2005), "person reliability" is equivalent to the traditional "test" reliability. Low values indicate a narrow range of person measures, or a small number of items. To increase person reliability, persons with more extreme abilities (high and low) should be tested, and the scale should be lengthened. The item reliability measure indicates how well items can be discriminated from one another based on their difficulty and has no traditional equivalent in classical test theory. Low values indicate a narrow range of item measures or a small sample and can likely be increased by testing more people (Linacre, 1991-2005).

Reliability for persons and items ranges from 0 to 1. The closer the reliability is to 1, the less the variability of the measurement can be attributed to measurement error. Winsteps computes upper and lower boundary values for the unknown, “true” reliability of persons

[2] While adequate mean square fit statistics are sufficient for assessing fit, an examination of the Zstd statistics reveals that they, too, are well within the range of -2.00 to 2.00, with values between -.2 and -.6.

and items. The lower boundary is the "real reliability." The upper boundary is the Model Reliability. Table 1 shows that, for the present data, the real item measure reliability is 1.00, and the person measure reliability is .96. These results indicate that the estimated measures are highly reliable (i.e., 0% and 4% respectively, of item and person measure variability, can be attributed to measurement error).

The index of separation was also examined as another measure of the fit of the data to the Rasch model. Separation refers to the spread of person positions or item positions along the variable, CSR implementation. If item separation is 1.0 or below, items may not have enough spread or breadth, and the items do not create a well-defined variable. Likewise, person separation indices less than 1.0 indicate that the scale does not discriminate well among respondents. As displayed in Table 1, the real item and person separation indices for the CSR Implementation Scale are 20.11 and 5.12, well above the lower limit of 1.0, indicating that items have sufficient breadth and persons are well-discriminated.

Unidimensionality

CSR Implementation Scale data were also examined to determine the degree to which they exhibited evidence of unidimensionality, a major assumption of the Rasch model. In other words, for data to fit the Rasch model, each item must contribute to the measurement of a single attribute (i.e., CSR implementation). One way of checking on this assumption is to perform a principal-components analysis of the observation residuals. The residuals are those parts of the observations *not explained* by the Rasch dimension. If large sub-structures are found, the scale developer should consider dividing the items into more than one instrument.

The principle components analysis of the residuals from the CSR Implementation Scale indicate that this instrument has a strong single dimension. In fact, the first component (i.e., the part of the observations *explained* by the Rasch model) account for 76% of the total variance. The second through sixth components account for a total of 8.8% of the total variance. An examination of the loadings on the second component indicated that the items relating to overall beliefs about CSR are located on the positive end of component's continuum, while items describing actions teachers and schools can take to implement CSR define the negative end of the continuum.

To further evaluate the dimensionality of the CSR Implementation Scale was split into two subtests, based on positive and negative loadings on the first residual factor (Linacre, 1991-2005, p. 248). The correlation between the two measures was computed, and all respondents were subsequently measured on the two subtests, and the two measures were cross-plotted (see Figure 2).

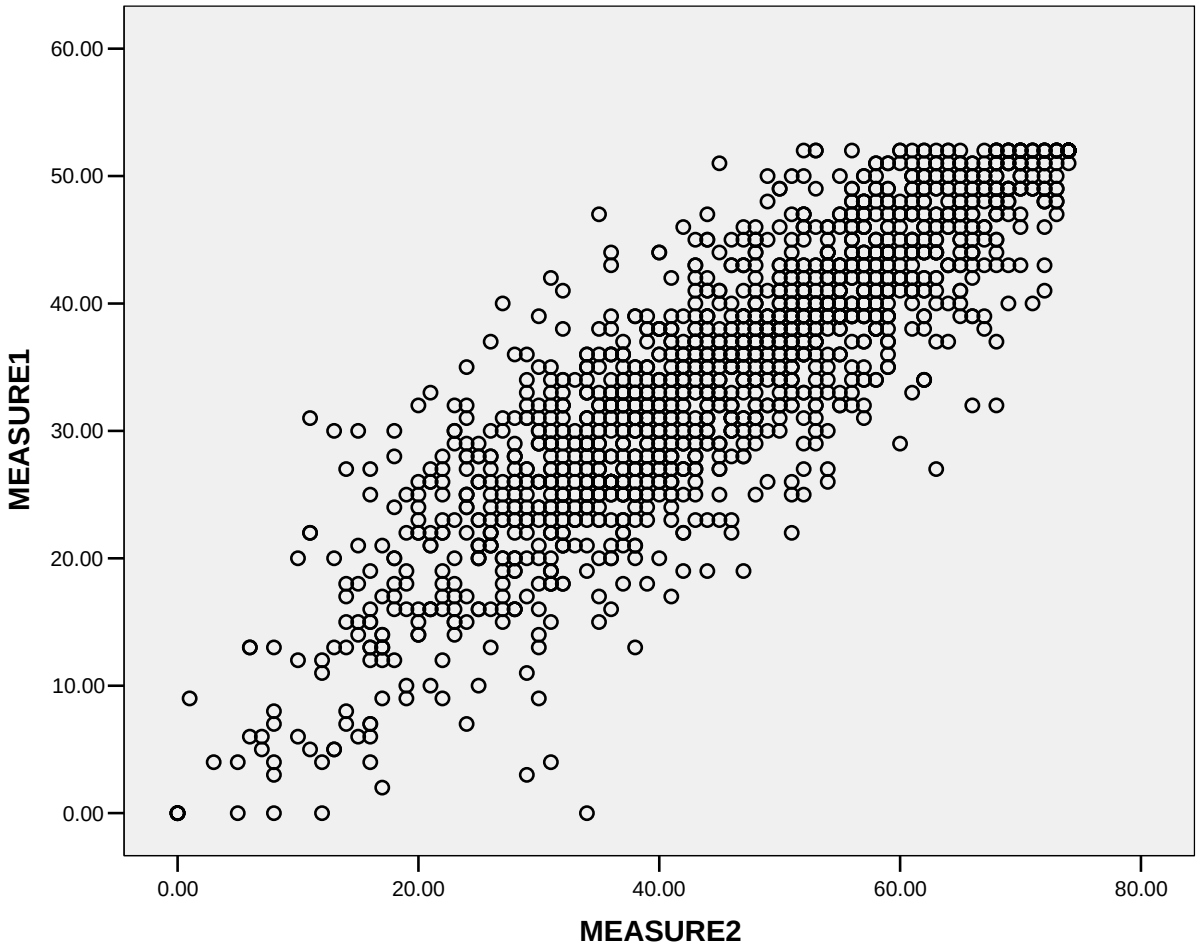

Figure 2: Correlation between Subtests

The correlation between the two measures was .88 ($p<.001$). A regression fit line was plotted to the data. Additionally, it was revealed that fewer than 5% of data points lay beyond ±2 standard deviations of the fit line. Using Linacre's guidelines: "If only a few people are noticeably off-diagonal, or off-diagonal deviance would not lead to any action, then you have a substantively unidimensional test" (Linacre, 1991-2005, p. 248), it was reasonable to conclude that the unidimensionality assumption was met by the CSR Implementation Scale. This judgment was further supported by the fact that the reliability of separation (based on empirical standard errors) of the CSR Implementation Scale is 1.00, indicating that the scale demonstrates good internal consistency.

Rating Category Effectiveness

Table 2 Response Scale Use and Figure 1 Category Probabilities display how the response scale was used. For these data, the response scale was 0 (strongly disagree) to 4 (strongly agree). Observed Count indicates the number of times the category was selected across all items and persons. Respondents were not likely to endorse a 0 (strongly disagree) category, with only 6% of responses in this category. The mean

squared estimates for each category, MNSQ, are always less than 1.4, indicating a lack of substantial misfit. The step calibration is expected to increase with category value, and it does.

Guidelines indicate that thresholds should increase by at least 1.4 logits, to show distinction between categories, but not more than 5 logits, in order to avoid large gaps in the variable (Linacre, 1999a). Table 2 below reveals that the CSR implementation rating scale does not quite meet this guideline, with the threshold distance between categories 1 and 2 just 0.94, suggesting that the distinction between step 1 and the midpoint of the scale is not as clear as it could be. In contrast, the thresholds between steps 2 and 3 and 3 and 4 are 1.75 and 1.8, respectively, well within the guidelines set forth for the magnitude of distances between threshold distances.

Table 2: Summary of Category Structure

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE	
0	0	3770	6	-1.67	-1.85	1.24	1.27	NONE	(-3.37)	0
1	1	8265	12	-.74	-.74	1.00	1.02	-2.05	-1.67	1
2	2	19547	29	.16	.23	.91	.94	-1.11	-.15	2
3	3	20945	31	1.32	1.31	.91	.85	.68	1.64	3
4	4	13235	20	2.90	2.85	1.07	1.07	2.48	(3.69)	4
MISSING		1246	2	.09						

AVERAGE MEASURE is mean of measures in category. It is not a parameter estimate.

CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----		50% CUM. PROBABLTY	COHERENCE M->C	C->M	ESTIM DISCR	
0	NONE		(-3.37)	-INF	-2.59		68%	22%		0
1	-2.05	.02	-1.67	-2.59	-.92	-2.32	41%	38%	.79	1
2	-1.11	.01	-.15	-.92	.70	-.99	52%	62%	.92	2
3	.68	.01	1.64	.70	2.80	.69	55%	67%	1.10	3
4	2.48	.01	(3.69)	2.80	+INF	2.62	81%	51%	1.08	4

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

Figure 3 below displays the probability curve for the CSR Implementation Scale, which are another useful means for examining the distinction between thresholds. Curves display the likelihood of category selection (y-axis) by the person-minus-item measure (x-axis). Each category should have a distinct peak in the probability curve graph, illustrating that it is the most probably response category for some portion of the variable, CSR implementation. The transition points between categories are the step calibration values from Table 2. If all categories are utilized, each category value will be the most likely at some point, and no curves will be inverted (i.e., there will be no points at which a higher category is more likely at a lower point than a lower category). A visual inspection of the probability curve in Figure 3 reveals that it meets the expectations

associated with it, except for the relatively small distinction between categories 1 and 2 mentioned earlier.

Figure 3 is useful for estimating the probability of endorsement of various categories of an item, given information related to the item difficulty and person ability. In the curve below, for a difference in logit position between a person and an item of +1, while any response is possible, the most likely category response is 3. Indeed, the probability of a response of 3 is approximately 45%, as opposed to 0% for a 0 response, about 5% for a 1 response, close to 10% for a 4 response, and approximately 40% for a 2 response.

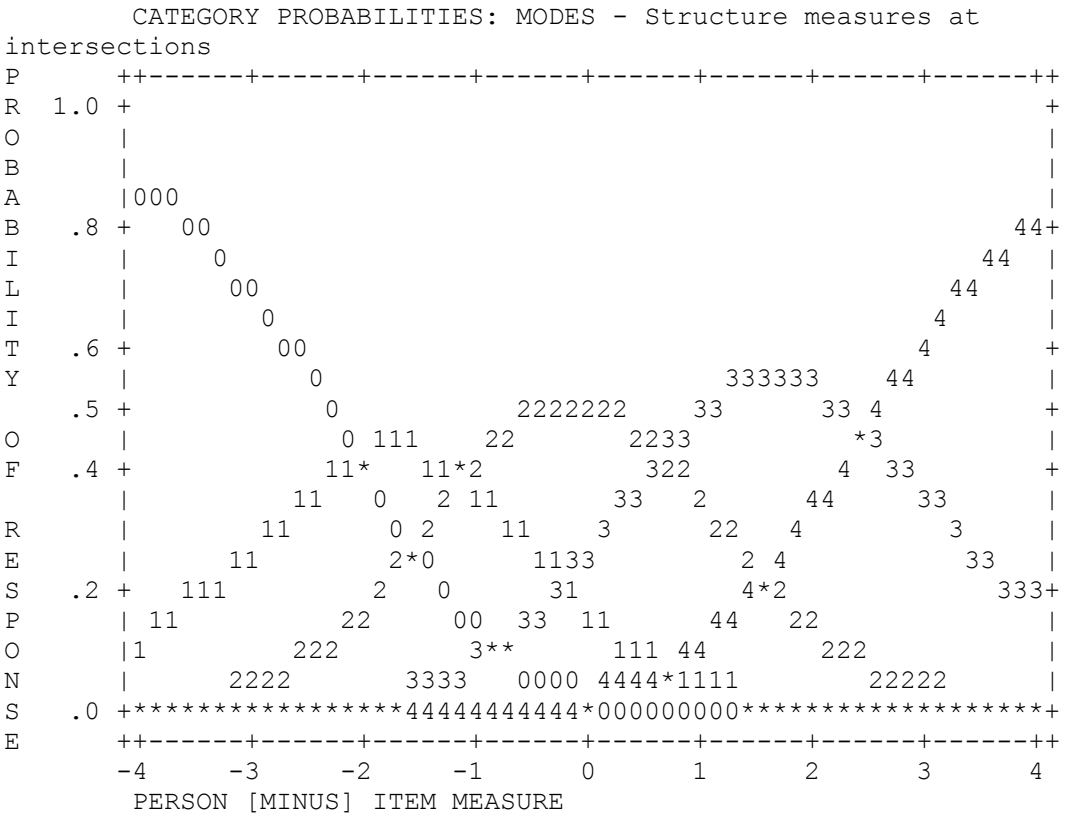

Figure 3: Category Probabilities

Knowing the person ability/implementation levels of any of the respondents of the CSR Implementation Scale, the probability of these persons agreeing with any of the items can be calculated, as described above. This is useful for being able to qualitatively describe the features of CSR implementation that any individual is likely to have put into place and to predict that person's response to any item.

Figure 4 also provides a useful illustration for predicting teacher's response to any CSR Implementation Scale item, given his/her ability score. For example, there is a 50% probability that any person with an ability/implementation score of 3 would assign a rating of 4 (strongly agree) to all of the items on the CSR Implementation Scale, with the exception of three items that would be more likely rated 3: QA4: Parents understand

and support CSR; AQ13: Community members understand and support CSR; QA37A: I have succeeded in changing my teaching; and QA37B: I have tried to change my teaching.

```
TABLE 2.3 C:\DOCUMENTS AND SETTINGS\SUSAN G REPORT OUTFILE.txt Apr  2 15:24 2005
INPUT: 2100 PERSONS, 32 ITEMS  MEASURED: 2099 PERSONS, 32 ITEMS, 5 CATS     3.55.0
--------------------------------------------------------------------------------

50% CUMULATIVE PROBABILITY: MEDIAN (ILLUSTRATED BY AN OBSERVED CATEGORY)
-5       -3        -1         1         3         5         7
|---------+---------+---------+---------+---------+---------|  NUM   ITEM
0                     1      2       3         4            4    8  QA4       PARENTS UNDERSTAND & SUPPORT CSR
0                    1      2       3         4             4   15  QA13      COMMUNITY UNDERSTANDS & SUPPORTS CSR
0                    1     2        3        4              4   32  NYTEA37B I HAVE CHANGED TCHG
0                   1     2        3        4               4   31  NYTEA37A I HAVE TRIED TO CHANGE TCHG
0              1      2       3         4                   4    9  QA5       DISTRICT REFORMS ALIGN W CSR
0              1      2       3         4                   4   24  QA29      SCHOOL PROVIDES ACADEMIC INFO TO FAMILIES
0              1      2       3         4                   4   16  QA15      CURRIC, INSTRUCT, MATLS VERTICALLY ALIGNED
0              1     2        3        4                    4   27  QA33      TCHRS CAN EXPRESS CONCERNS TO ADMINISTRATION
0             1      2       3         4                   4   10  QA6       DISTRICT ADMINISTRATORS SUPPORT CSR
0             1      2       3         4                   4   30  QA35G     PD HELPED ME USE RESEARCH-BASED PRACTICES
0             1     2        3        4                    4   19  QA22      TEACHERS ASSESS CSR PROGRESS
0             1     2        3        4                    4   25  QA31      I RECEIVE SUPPORT TO IMPLEMENT CSR
0            1      2       3         4                   4    3  QA1C      CSR HELPS TCHRS MEET SS NEEDS
0            1      2       3         4                   4   17  QA16      CURRIC, INSTRUCT, MATLS HORIZONTALLY ALIGNED
0            1      2       3         4                   4   26  QA32      TCHRS HAVE OPPORTUNITY TO DISCUSS CSR
0            1      2       3         4                   4    7  QA2       TCHRS SUPPORT CSR
0            1     2        3        4                    4   29  QA35F     PD INCREASED MY UNDERSTANDING OF CONTENT
0           1      2       3         4                   4   18  QA21      COLLEAGUES ARE TEAM PLAYERS
0           1      2       3         4                   4    6  QA1G      CSR BASED ON RESEARCH, BEST PRACTICE
0           1      2       3         4                   4   23  QA26      SCHOOL MAKES DATA-DRIVEN DECISIONS
0           1      2       3         4                   4    4  QA1D      CSR INVOLVES WHOLE SCHOOL
0           1      2       3         4                   4   12  QA8       CSR LEADERS WORK WELL TOGETHER
0           1      2       3         4                   4   22  QA25      ADMINISTRATORS & TCHRS SHARE CSR RESPONSIBILITY
0           1      2       3         4                   4   28  QA35E     PD GIVES ME TCHG IDEAS
0           1      2       3         4                   4   11  QA7       LDSHP TEAM ACTIVE IN PROMOTING CSR
0          1      2       3         4                    4    1  QA1A      CSR HELPS SS MEET STDS
0          1      2       3         4                    4    2  QA1B      CSR HELPS TCHRS ALIGN C,I,A W STDS
0          1     2        3        4                     4    5  QA1F      CSR FOCUSED ON IMPROVING TCHG, LRNG
0         1      2       3         4                    4   13  QA9       I SUPPORT SCHOOL CSR APPROACH
0         1      2       3         4                    4   21  QA24      I UNDERSTAND RELAT. BTWN CSR & SCHOOL GOALS
0       1      2       3         4                      4   20  QA23      I UNDERSTAND SCHOOL GOALS
0       1     2        3        4                       4   14  QA10      I AM COMMITTED TO TCHG CHANGES
|---------+---------+---------+---------+---------+---------|  NUM   ITEM
-5       -3        -1         1         3         5         7

                        111 11 111
              11 22245789211810800076645323122111    2
5  1 3 233555019774542487313609084886427819117319 27 8  3       PERSONS
               T        S       M        S        T
```

Figure 4: Most Probable Responses/Scores

Variable Map

The variable map is a graphic that reveals the operational definition of what the CSR Implementation Scale is measuring. Teachers are to the left of the vertical line, and items are to the right. At the top of the map are teachers with the highest CSR implementation ratings and the most difficult items (i.e., those requiring the highest levels of CSR implementation for agreement with the item). At the bottom of the map are low CSR implementing teachers and the "easier" items (i.e., those requiring low levels of CSR implementation for agreement). The variable map allows one to see how items spread out over the continuum of CSR implementation and how teachers distribute on the CSR implementation variable.

Items should be located at each point on the scale to measure meaningful differences in CSR implementation. The items should cover all areas of the ruler to measure the CSR implementation of all teachers who respond to the CSR Implementation Scale. The variable map for the CSR Implementation Scale is presented in Figure 5.

An examination of the variable map in Figure 5 reveals substantial gaps in the continuum represented by the CSR Implementation Scale. While person ability spans nearly 8 logits, items span less than 3.5 logits. In fact, the person distribution is top heavy compared to the item distribution, revealing that the CSR Implementation Scale is a relatively "easy" scale (i.e., it does not adequately tap into truly "high" levels of CSR implementation). Given the fact that some of scale respondents had been implementing CSR for as many as five years, a significant number of them can be considered high implementers and are not well-targeted by the CSR Implementation Scale, as revealed by a significant gap in the coverage of the construct at the upper end of the scale where there are people but no items. While the scale reveals that there is a considerable group of high implementers, it is not able to characterize the practices associated with high levels of implementation. Considering experienced, high CSR implementers would be part of any target group for a scale such as this, the scale needs more difficult items so that the CSR implementation levels of teachers at the top of the continuum can be more precisely estimated. The scale would also be improved by the addition of two to three items at the lower, "easier" end of the scale, where there is a small area without any items tapping into the CSR implementation levels of very low implementers.

The variable map in Figure 5 also reveals considerable redundancy in the items making up the CSR Implementation Scale. For example, six items measure CSR implementation features at the difficulty level corresponding to a logit value of 0 (the mean of the scale). Five of the six items could be eliminated, and the scale would still accurately measure the CSR implementation level of teachers with implementation levels at the mean of the scale. Furthermore, 20 of the 32 items in the CSR Implementation Scale (63%) measure implementation at difficulty levels between 0 and -1 logits. The scale would be significantly improved by deleting many of these redundant items and adding items that discriminate more effectively along more points on the scale

```
.
TABLE 1.2 C:\DOCUMENTS AND SETTINGS\SUSAN G REPORT OUTFILE.txt Apr  2 15:24 2005
INPUT: 2100 PERSONS, 32 ITEMS  MEASURED: 2099 PERSONS, 32 ITEMS
--------------------------------------------------------------------------------

       PERSONS MAP OF ITEMS
               <more>|<rare>
    6             .  +
                     |
                     |
                .##  |
                     |
    5             .  +
                  .  |
                  .  |
                 .#  |
                 .#  |
    4            .# T+
                .##  |
                .##  |
                 .#  |
               .###  |
    3           .##  +
               .###  |
             .#####  |
              .#### S|
            .######  |
    2       .######  +
          .#######  |  QA4
       .##########  |  QA13
       .##########  |  QA37B
        ##########  |T
    1     .########  +  QA37A
        ########## M|
      .###########  |S
        .########  |  QA15   QA29   QA5
      .###########  |  QA33   QA35G  QA6
    0  .###########  +M QA16   QA1C   QA2    QA22   QA31   QA32
     .############  |  QA1D   QA1G   QA21   QA25   QA26   QA35E  QA35F  QA8
        .#########  |  QA1A   QA1B   QA7
         .########  |S QA1F   QA24   QA9
          .####### S|
   -1       .#####  +  QA23
             .####  |T QA10
               .##  |
               .##  |
               .##  |
   -2            .  +
                .#  |
                 # T|
                 .  |
                 .  |
   -3            .  +
                 .  |
                 .  |
                 .  |
                    |
   -4            .  +
                    |
                 .  |
                    |
                    |
   -5            .  +
              <less>|<frequ>
 EACH '#' IS 10.
```

Figure 5: Variable Map with Item Labels

Figure 6 displays a variable map containing the actual wording of the items on the CSR Implementation Scale. This representation of the map is useful for examining the nature of CSR implementation and what characterizes difficulty along the continuum of CSR implementation. For example, the variable map shows that the aspects of CSR implementation that teachers find most difficult have to do with the Parent and Community Involvement component of CSR. Even if teachers and schools support CSR and have instituted significant instructional changes ("easy" items at the opposite end of the scale), it is not likely that parents and the community understand and support CSR. The next most difficult aspects of CSR implementation involve changing teaching practice. Even when many other features of CSR are in place, changing what goes on in the classroom is a much more difficult task. While these findings are highly intuitive (i.e., changing practice is more difficult that understanding or supporting an initiative), these results provide empirical evidence of this distinction.

Interestingly, the "easiest" item on the scale is "I am committed to maintaining the changes I have made in my teaching," reflecting a disconnect between the intentions to change practice at the "easy" end of the scale and the reality of changing instructional practice at the difficult end of the scale. Other items at the low end of the scale emphasize teachers' understanding and support of CSR and school goals. This suggests that intentions and attitudes related to CSR are implemented or made real before action steps related to actual implementation.

Data in the variable map also suggest that district and administrative level features of CSR implementation and alignment of curriculum, instruction, and materials are likely to be found after teachers have formulated positive intentions and attitudes toward CSR, yet changes in teaching and parent/community understanding and support has occurred.

An examination of the variable map also revealed interesting information about steps in CSR implementation that were at the same level of difficulty. For example, QA16, QA1C, QA2, QA22, QA31, AND QA32 all have the same logit difficulty value of zero. This suggests that teacher discussions of and assessment of the progress of CSR, teacher support of and for CSR, and horizontal alignment of curriculum, instruction, and materials are all likely to occur at the same point in implementation. The same is true of several other sets of items in the scale.

```
TABLE 12.2 C:\DOCUMENTS AND SETTINGS\SUSAN GRACIA ZOU027ws.txt Apr  2 14:52 2005
INPUT: 2100 PERSONS, 32 ITEMS  MEASURED: 2099 PERSONS, 32 ITEMS, 5 CATS     3.55.0
--------------------------------------------------------------------------------

       PERSONS MAP OF ITEMS
               <more>|<rare>
    6              .  +
                      |
                      |
                 .##  |
                      |
    5              .  +
                   .  |
                   .  |
                  .#  |
                  .#  |                                                .
    4             .# T+
                  .##  |
                  .##  |
                   .#  |
                 .###  |
    3             .##  +
                 .###  |
              .######  |
                .#### S|
             .######  |
    2        .######  +
           .#######  |  QA4       PARENTS UNDERSTAND & SUPPORT CSR
        .###########  |  QA13      COMMUNITY UNDERSTANDS & SUPPORTS CSR
        .###########  |  QA37B     I HAVE CHANGED TCHG
         ##########  |T
    1     .########  +  QA37A     I HAVE TRIED TO CHANGE TCHG
         ########## M|
        .###########  |S
           .########  |  QA15      CURRIC, INSTRUCT, MATLS VERTICALLY ALIGNED
                         QA29      SCHOOL PROVIDES ACADEMIC INFO TO FAMILIES
                         QA5       DISTRICT REFORMS ALIGN W CSR
        .###########  |  QA33      TCHRS CAN EXPRESS CONCERNS TO ADMINISTRATION
                         QA35G     PD HELPED ME USE RESEARCH-BASED PRACTICES
                         QA6       DISTRICT ADMINISTRATORS SUPPORT CSR
    0  .###########  +M QA16      CURRIC, INSTRUCT, MATLS HORIZONTALLY ALIGNED
                         QA1C      CSR HELPS TCHRS MEET SS NEEDS
                         QA2       TCHRS SUPPORT CSR
                         QA22      TEACHERS ASSESS CSR PROGRESS
                         QA31      I RECEIVE SUPPORT TO IMPLEMENT CSR
                         QA32      TCHRS HAVE OPPORTUNITY TO DISCUSS CSR
      .############  |  QA1D      CSR INVOLVES WHOLE SCHOOL
                         QA1G      CSR BASED ON RESEARCH, BEST PRACTICE
                         QA21      COLLEAGUES ARE TEAM PLAYERS
                         QA25      ADMINISTRATORS & TCHRS SHARE CSR RESPONSIBILITY
                         QA26      SCHOOL MAKES DATA-DRIVEN DECISIONS
                         QA35E     PD GIVES ME TCHG IDEAS
                         QA35F     PD INCREASED MY UNDERSTANDING OF CONTENT
                         QA8       CSR LEADERS WORK WELL TOGETHER
        .#########  |  QA1A      CSR HELPS SS MEET STDS
                         QA1B      CSR HELPS TCHRS ALIGN C,I,A W STDS
                         QA7       LDSHP TEAM ACTIVE IN PROMOTING CSR
         .########  |S QA1F      CSR FOCUSED ON IMPROVING TCHG, LRNG
                         QA24      I UNDERSTAND RELAT. BTWN CSR & SCHOOL GOALS
                         QA9       I SUPPORT SCHOOL CSR APPROACH
          .####### S|
   -1       .#####  +  QA23      I UNDERSTAND SCHOOL GOALS
              .####  |T QA10      I AM COMMITTED TO TCHG CHANGES
                .##  |
                .##  |
                .##  |
   -2             .  +
                 .#  |
                  # T|
                  .  |
                  .  |
   -3             .  +
                  .  |
                  .  |
                  .  |
                     |
   -4             .  +
                     |
                  .  |
                     |
                     |
   -5             .  +
               <less>|<frequ>
 EACH '#' IS 10.
```

Figure 6: Variable Map with Item Names

Discussion

Looking at fit statistics from initial Winsteps output, it was evident that twenty-one items in the original CSR Implementation Scale were measuring a different construct from the other 32 items. These items were eliminated, and analyses continued with a reduced CSR Implementation Scale.

Analyses revealed that the 32 item CSR Implementation Scale fits the Rasch model. The mean square infit and outfit statistics for persons are 1.02 and 1.00, well within the .75-1.3 range recommended by Bond and Fox (2001). Similarly, the mean square infit and outfit statistics for items are 1.01 and 1.00. In each case, the expected mean square value of 1.00 is achieved. Furthermore, the CSR Implementation Scale met the assumption of unidimensionality. This provides evidence of the validity of the CSR Implementation Scale, in that the items worked as required together to measure a single variable, the items form a hierarchy, and teachers' response patterns were acceptable given the expected hierarchy of responses (Wright and Stone, 1979).

The real item measure reliability for the CSR Implementation Scale was 1.00, and the person measure reliability was .96. These results indicate that the estimated measures are highly reliable (i.e., 0% and 4% respectively, of item and person measure variability, can be attributed to measurement error). The real item and person separation indices for the CSR Implementation Scale are 20.11 and 5.12, well above the lower limit of 1.0, indicating that items have sufficient breadth and persons are well-discriminated.

The mean squared estimates for each rating scale category, MNSQ, were always less than 1.4, indicating a lack of substantial misfit. The step calibration wass expected to increase with category value, and it did. However, the threshold distance between categories 1 and 2 on the five-point scale was just 0.94, suggesting that the distinction between step 1 and the midpoint of the scale is not as clear as it could be.

The variable map revealed a hierarchy of CSR implementation indicators. Items at the low end of the scale emphasize teachers' understanding and support of CSR and school goals. This suggests that intentions and attitudes related to CSR are implemented or made real before action steps related to actual implementation. Data in the variable map also suggest that district and administrative level features of CSR implementation and alignment of curriculum, instruction, and materials are likely to be found after teachers have formulated positive intentions and attitudes toward CSR.

The aspects of CSR implementation that teachers find most difficult have to do with the Parent and Community Involvement component of CSR. Even if teachers and schools support CSR and have instituted significant instructional changes ("easy" items at the opposite end of the scale), it is not likely that parents and the community understand and support CSR. The next most difficult aspects of CSR implementation involve changing teaching practice. Even when many other features of CSR are in place, changing what goes on in the classroom is a much more difficult task. While these findings are highly

intuitive (i.e., changing practice is more difficult that understanding or supporting an initiative), these results provide empirical evidence of this distinction.

Nevertheless, the variable map revealed substantial gaps in the continuum represented by the CSR Implementation Scale. While person ability spanned nearly 8 logits, items spanned less than 3.5 logits. In fact, the person distribution was top heavy compared to the item distribution, revealing that the CSR Implementation Scale is a relatively "easy" scale (i.e., it does not adequately tap into truly "high" levels of CSR implementation). Given the fact that a significant percentage of scale respondents had been implementing CSR for more than three years, a significant number of them can be considered high implementers and are not well-targeted by the CSR Implementation Scale, as revealed by a significant gap in the coverage of the construct at the upper end of the scale where there are people but no items. While the scale revealed that there is a considerable group of high implementers, it is not able to characterize the practices associated with high levels of implementation. Considering experienced, high CSR implementers would be part of any target group for a scale such as this, the scale needs more difficult items so that the CSR implementation levels of teachers at the top of the continuum can be more precisely estimated. The scale would also be improved by the addition of two to three items at the lower, "easier" end of the scale, where there is a small area without any items tapping into the CSR implementation levels of very low implementers.

The variable map also revealed considerable redundancy in the items making up the CSR Implementation Scale. For example, six items measure CSR implementation features at the difficulty level corresponding to a logit value of 0 (the mean of the scale). Five of the six items could be eliminated, and the scale would still accurately measure the CSR implementation level of teachers with implementation levels at the mean of the scale. Furthermore, 20 of the 32 items in the CSR Implementation Scale (63%) measure implementation at difficulty levels between 0 and -1 logits. The scale would be significantly improved by deleting many of these redundant items and adding items that discriminate more effectively along more points on the scale

References

Berman, P. & McLaughlin, M. (1977). *Federal programs supporting educational change: Vol. 7. Factors affecting implementation and continuation*. Santa Monica, CA: Rand Corporation.

Berman, P., & McLaughlin, M. W. (1978). *Federal programs supporting educational change, Vol. VIII: Implementing and sustaining innovations*. Santa Monica, CA: Rand

Bond, T.G. & Fox, C. (2001). *Applying the Rasch Model*. New Jersey: Lawrence Erlbaum.

Education Commission of the States. (1999). A promising approach for today's schools. *Comprehensive School Reform, 1* (3). Available: http://www.ecs.org/clearinghouse/16/42/1642.doc.

Elmore, R. F. (1996). Getting to scale with successful educational practices. *Harvard Educational Review, 66*(1), 1–26

Linacre, J. M. (2005) *WINSTEPS Rasch Measurement Computer Program*. Chicago: Winsteps.com.

Linacre, J.M. (1991-2005). *A User's Guide to Winsteps/Ministeps Rasch-Model Computer Programs.* Chicago, IL: Winsteps.

Little, J.W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American Educational Research Journal, 19*, 325-340.

Sampson, S.O. and Bradley, K.D. (2003). Rasch analysis of educator supply and demand rating scale data: An alternative to the true score model. *Research Methods Forum.* Available at: http://division.aomonline.org/rm/2003forum/rasch.pdf

Smith, E., Jr. (2000). *Rasch Measurement Models*. Paper presented at An Introduction to Rasch Measurement: Theory and Applications, Chicago.

U.S. Department of Education, (2002). *Guidance on the Comprehensive School Reform Program*. Washington, DC: U.S. Department of Education. Available at: http://www.ed.gov/programs/compreform/guidance/index.html

Wright, B. D. & Stone, M. H. (1979). *Best Test Design*. Chicago: MESA Press.