Statistical Consulting Laboratory
Department of Statistics
University of South Carolina

Statistical Analysis of the South Carolina Report Card Surveys

FINAL REPORT

Awarding Agency:
The State of South Carolina
Education Oversight Committee

June 2005

Research Team

Roumen Vesselinov, Ph.D.

John Grego, Ph.D.

Brian Habing, Ph.D.

Emmy McKinney, M.A.

Lalita Das, Ph.D. student

Questions can be directed to:

Roumen Vesselinov
Department of Statistics
University of South Carolina
LeConte College 200-C
Columbia, SC 29208

(803) 777-5074 Phone
(803) 777-4048 Fax
roumen@sc.edu

Acknowledgements

The authors would like to thank the Department of Statistics at USC, particularly Dr. James Lynch, Chair of the Department, and Debra Williams, Business Manager of the Department.

Thanks should be extended also to the USC Sponsored Awards Management (SAM) and particularly Ms. Danielle McElwain and Mr. R. Steven Etheredge, Director of Pre-Award Services.

We would like to thank Mr. David Potter, Director of Research and the whole Education Oversight Committee for the opportunity to work on this interesting and challenging project.

CONTENTS

INTRODUCTION

This project was initiated by Mr. David Potter, Director of Research for the SC Education Oversight Committee (EOC). He approached the Statistical Consulting Laboratory in April of 2005.

The main objective of this project is to test the quality of the scales used in the SC Report Card Surveys. We were supplied with the data for the 2003/4 school year. Originally we planned to analyze the scales using combined data from all schools and by the school type: elementary, middle and high school. Later on, during the initial phase we discovered that there were 3 additional types of schools: career centers, primary schools, and special schools. These 3 types were added to the analysis. See the scales and their items in the next section of this report.

Working in close cooperation with Mr. David Potter we were able to identify the school type for each school which in some cases proved difficult. Finally we arrived at the following samples and surveys.

Table I1. Sample Size

School Type	Student Surveys	Teacher Surveys	Parent Surveys
1. Elementary Schools	52,138	20,571	33,780
2 Middle Schools	43,829	9,010	20,124
3. High Schools	32,471	10,881	10,890
4. Career Centers	3,615	680	1,299
5. Primary Schools		825	
6. Special Schools	483	148	208
Total	**132,536**	**42,115**	**66,301**

Primary schools had data only from teachers and special schools had relatively few observations.

The student and teacher surveys have 3 subscales and one overall scale. The parent survey has 7 subscales and one overall scale. There are 16 scales in total for testing with combined data from all schools and by school type. Overall, analysis was performed on 16 scales with 19 different samples, or 304 scale analyses altogether. Each scale analysis involves 4 different groups of methods. For details on the methods see the Methodology section of this report.

Obviously this large number of analyses, samples and school types was one of the challenges of this project. We tried to present the results in as concise format as possible. For the combined data analysis the results for each subscale are presented in a summary table. For the analysis by school type the results are presented in one summary table per scale where the accent is on the possible differences by school type and combined data from all schools. For both types of analysis, with combined data and by school type, two separate appendices contain all the summary information from the 4 different methods for each scale. These appendices can be useful for learning more details than the body of the text contains.

ACRONYMS USED IN THE TEXT

FA	–	Factor Analysis
RC	–	Reproducibility and Consistency
CC	–	Conditional Covariance based methods
HCA	–	Hierarchical Cluster Analysis
MDS	–	Multidimensional Scaling
LE	–	Learning Environment
SPE	–	Social and Physical Environment
HSR	–	Home-School Relations
H	–	Loevinger's H coefficient
MO	–	Mokken scale

METHODOLOGY

Scales – Basic Concepts

Scales can be described as ordinal indices that are thought to measure a latent variable. In a narrow sense scales are defined as sets of items which stand in ordinal relationship to each other. **Guttman Scales** variables with two response levels consist of set of items constituting a unidimensional series such that an answer to a given item predicts the answers to all previous items in the series. This is a deterministic scale which means that if a respondent answers positively to a particular item he/she will answer all less difficult items positively too. Guttman scales for polytomous items have similar characteristics.

Difficulty of an item is related to the propensity to endorse a particular item. For example, the item "*Does your school do **everything possible** to help students?*" has low propensity of endorsement, i.e. it is a more difficult item. The item "*Do you think your school **at least tries** to help students?*" has higher propensity for endorsement, i.e. it is less difficult.

Mokken Scales are similar to Guttman scales but they are probabilistic, i.e. in Mokken scales a respondent answering an item positively will have a significantly greater probability of answering a less difficult item in a positive way as well.

Likert Scales are scales for which the ordinality refers only to an ordinal relationship of values within a single item. These values are not necessarily ordinal with respect to each other (item to item).

Under the narrow definition of a scale only Guttman and Mokken scales are real scales. Likert scales do not necessarily meet this strict criterion although sets of Likert items can be used to form scales.

In a Mokken scale all items have different difficulties that are reflected in different proportions of positive responses. The trace line of an item is a plot of the probability of positive response to an item (Y axis) against the value of a latent trait (X axis). The trace lines of items should be monotone and should not intersect, i.e. there is no double monotonicity. In addition, the trace lines should be steep enough to produce only a small number of Guttman errors. A **Guttman error** is an exception to the rule that a positive answer to one item implies a positive answer to a less difficult item.

Loevinger's H coefficient measures the conformity of a set of items to Mokken criteria and validates their use together as a scale of a unidimensional latent variable. It is based on the

number of observed Guttman errors and the total errors expected under the Null hypothesis that items are totally unrelated. When there are no Guttman errors, H_{ij} =1(item j is easier that item i). When the response is random H_{ij}=0. One can sum across all j's to get H_i which gives a measure of the fit of item "i" to the Mokken scale.

If we average across all item pairs we get the overall H which is a measure for the whole scale. We will consider the following **classification of scales** corresponding to the overall Loevinger's H coefficient:

1. Weak scale: $0.3 \leq H < 0.4$
2. Moderate scale: $0.4 \leq H < 0.5$
3. Strong scale: $H \geq 0.5$

The threshold of H=0.3 is arbitrary but widely used for practical purposes. Cases with H<0.3 cannot be considered scales in a Mokken sense.

Assumption of unidimensionality. Scales are usually assumed to measure a single trait, dimension or meaning.

METHODS:

A. <u>Factor Analysis (FA)</u>

Factor analysis can be described as a family of techniques for investigation of the relationship between a set of observed variables and underlying latent factors or dimensions. The observed variables can be thought of as components of the latent variables. The model is trying to explain the variance in the observed variables in regard to the latent factors. Linear FA is analogous to multivariate linear regression model where the dependent variable is the latent trait, and the independent variables are the observed variables, while the factor loadings play the role of regression coefficients.

If a group of variables is shown to relate strongly to the same dimension, it can be said that they share this dimension in common. By determining what variables relate to a given dimension and defining what is common to them and how they differ from the other unrelated variables we can derive an understanding of this dimension.

In theory, linear FA is designed for continuous response variables, although it generally works well for items with 5 or more categories. For 3 or 4 category items linear FA can be used with caution. There are some difficulties that can occur for 2 category items so in general linear FA is not applied here.

There are two basic approaches to factor analysis: exploratory and confirmatory factor analysis. Exploratory factor analysis tries to uncover the underlying dimensions. Confirmatory factor analysis is based on a theoretical model of the dimensions and their underlying components and it tries to determine to what extent the dimensions explain the relationships among observed variables. We will be employing exploratory factor analysis in this project.

Initial Extraction.

There are many different methods for extraction of the factors, including principal component FA, principal factor FA, maximum likelihood FA, iterated principal factor FA, image factoring FA, unweighted and generalized least squares, etc. In practice, very often different methods for extraction produce the same results. We have chosen maximum likelihood factor analysis for this project. The data is assumed to be multivariate normal but the method is somewhat robust to assumption violations and it is very flexible.

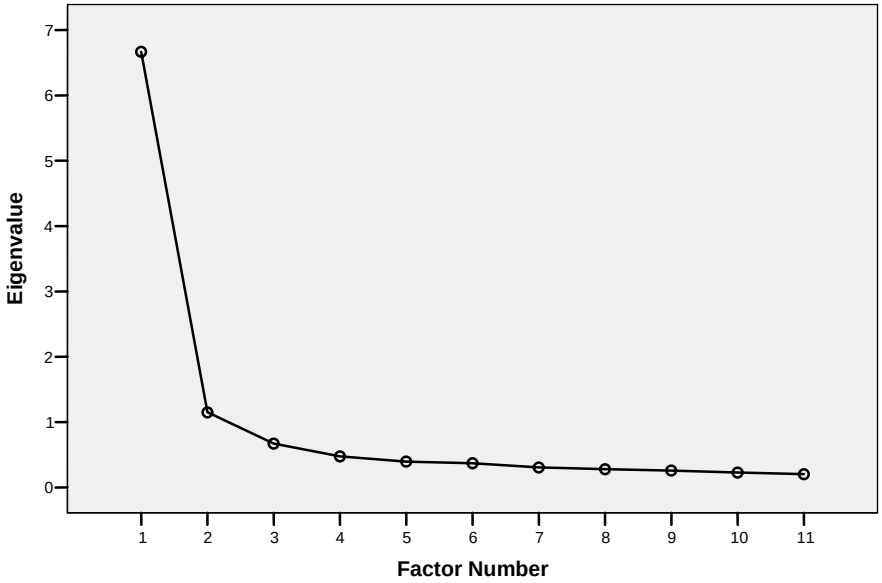
The rule of thumb for the sample size is that we should have at least 50 observations and at least 5 times as many observations as variables. Our samples satisfy this criterion. The loadings on the factors will use a threshold of loading greater than 0.3. In other words, if the coefficient for a particular item is greater than 0.3 this items will be considered to be part of this factor. In some cases one item may be highly correlated (>0.3) to more than one factor. It this case we will note that this item is double loaded.

Number of Factors.

For selecting the number of factors we will use a combination of rules:

1. **Eigenvalues greater than one** rule - all factors with eigenvalue greater than one will be possible candidates for inclusion.
2. The **scree plot** represents the number of possible factors versus the eigenvalues. It will give us additional guidance as to where the effect levels off or the "elbow effect". We will take the highest number of factors before the effect levels off. A typical scree plot is shown in Figure 1.
3. The **percent explained variance** will give us additional reason for including or excluding a factor**.** It seems reasonable that a good model should be able to explain at least 50% of the total variance. If a factor explains insignificant amount of the variance it can be excluded from analysis.

Figure 1. Scree Plot



From the initial phase the communalities for each variable are used. They show the proportion of the variance for that variable accounted for by the factors. If the communality for a particular variable is low (e.g., <.50) this would mean that the factor analysis does not account for a lot of this variable's variance. The reason could be that this variable is very different from the other variables and measures something different, or the measurement of this variable is very unreliable, or that the number of factors extracted is too small.

Factor Rotation.

After the initial extraction, the axes (factors) are rotated so they will more closely align with the points representing the variables. The rotations can be two types, orthogonal and oblique. The group of orthogonal rotations includes Varimax, Quartimax, Equamax, etc. The oblique group includes Direct Oblimin, Promax, and Orthoblique. For orthogonal rotation, the factors are kept independent from each other and their interpretation is very clear. Oblique rotations allow the factors to be correlated with each other which makes their interpretation more difficult but the assumptions are more realistic. In this analysis we will use Varimax rotation to determine the factors and Promax rotation for additional consideration and confirmation of the results.

Table 1. Communalities*

	Initial	Extraction
Q43	.662	.806
Q44	.657	.758
Q45	.600	.639
Q46	.663	.660
Q47	.674	.638
Q48	.573	.596
Q49	.600	.591
Q50	.671	.709
Q51	.632	.637
Q52	.547	.551
Q53	.725	.775

*Extraction Method: Maximum Likelihood

Example:

Table 1 above shows the communalities for each variable included in FA. None of them is less than 0.5 but if there were any we may consider removing these items or further investigating them particularly in the factor loadings in the rotated matrix. The example in Table 2 using the Varimax rotation shows that Factor #1 is highly related to items Q53, Q50, Q46, Q48, Q51, Q49, Q47, and Q52, and Factor #2 is highly related to items Q44, Q43 and Q45. Three items have significant loadings on both factors: Q45, Q46, and Q47.

Table 2. Rotated Factor Matrix*

	Factor	
	1	2
Q53	.762	.297
Q50	.726	.285
Q46	.706	.355
Q48	.701	.178
Q51	.699	.205
Q49	.696	.227
Q47	.653	.399
Q52	.639	.289
Q44	.249	.873
Q43	.264	.824
Q45	.393	.702

* Extraction Method: Maximum Likelihood.
Rotation Method: Varimax with Kaiser Normalization.

Table 3. Pattern Factor Matrix*

	Factor	
	1	2
Q53	.802	.024
Q48	.780	-.094
Q51	.766	-.061
Q50	.763	.026
Q49	.752	-.032
Q46	.708	.119
Q52	.655	.069
Q47	.624	.198
Q44	-.078	.956
Q43	-.039	.890
Q45	.172	.683

* Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.

In Table 3 the pattern matrix from a Promax rotation of the same factors is presented. In this case it basically confirms the findings of the Varimax rotation.

The model explains 61.7% of the total variance, with Factor 1 explaining 38.1% and Factor 2 explaining 23.6%.

B. Reproducibility and Consistency (RC)

Cronbach's Alpha is the most common estimate of reproducibility and consistency of a scale. This coefficient measures the extent to which item responses correlate highly with each other. Cronbach's alpha is not a measure of unidimensionality. For example an instrument with two very distinctive parts with highly correlated items within each part can get very high Cronbach's alpha regardless of the obvious two dimensions. Separate dimensions (clusters) can be highly correlated internally but not correlated between each other. Alpha is an estimated lower bound on reliability.

In general reliability will increase with the number of items, even when the set of items is not unidimensional. Reliability is a necessary but not sufficient condition for validity of scale items.

We will use the widely accepted cut-off of Alpha=0.70 or higher for a set of items to be considered a reliable scale. When Alpha is at least 0.70 the standard error of measurement will be less than 0.55 standard deviations. In addition to Cronbach's Alpha we will use two other measures for reliability, Rho and Lambda whose interpretation is very similar to the one for Cronbach's Alpha.

Example. Table 4A and Table 4B below show the Cronbach's Alpha for a scale of 11 items to be equal to 0.937 which means that the scale has very high reproducibility and consistency. The last column shows the scale's Alpha if a particular item is removed. For this example the reliability goes down if any item is excluded. If there was an item which when excluded increases the scale reliability or does not change the scale we can suggest removing this item from the scale.

Table 4A. Reproducibility and Consistency

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.937	.938	11

Table 4B. Item-Total Statistics

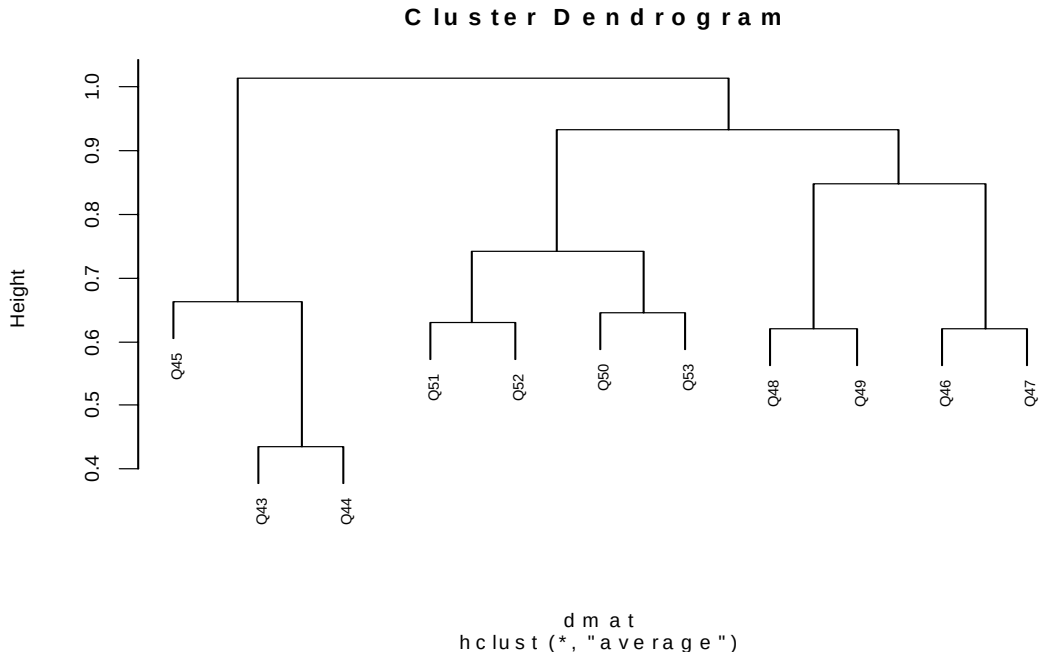
Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	**Cronbach's Alpha if Item Deleted**
Q43	31.02	41.172	.630	.667	.936
Q44	30.98	41.337	.641	.662	.935
Q45	31.27	39.688	.711	.604	.932
Q46	31.57	38.692	.779	.670	.930
Q47	31.41	39.069	.779	.680	.930
Q48	31.47	39.513	.720	.579	.932
Q49	31.47	39.662	.733	.605	.932
Q50	31.66	38.024	.792	.677	.929
Q51	31.75	37.134	.749	.636	.932
Q52	31.34	38.682	.722	.547	.932
Q53	31.61	36.836	.835	.730	.927

C. Conditional Covariance (CC) - Based Nonparametric Multidimensionality Assessment

Step 1. Hierarchical Cluster Analysis (HCA)

We will be using Conditional Covariance-Based Nonparametric multidimensionality assessment (Stout et al, 1996). The procedure is available through the software package HCA/CCPROX which stands for agglomerative hierarchical cluster analysis. The package performs a latent multidimensionality-sensitive hierarchical cluster analysis on either dichotomous or polytomous items. This nonparametric procedure is able to quickly cluster the items into progressively larger and larger dimensionally homogeneous groups. It allows us to examine the scale's dimensionality at a variety of agglomeration levels, ranging from which pairs of items are most closely dimensionally related, to which two item clusters best dimensionally summarize the entire scale.

Figure 2. Hierarchical Cluster Analysis



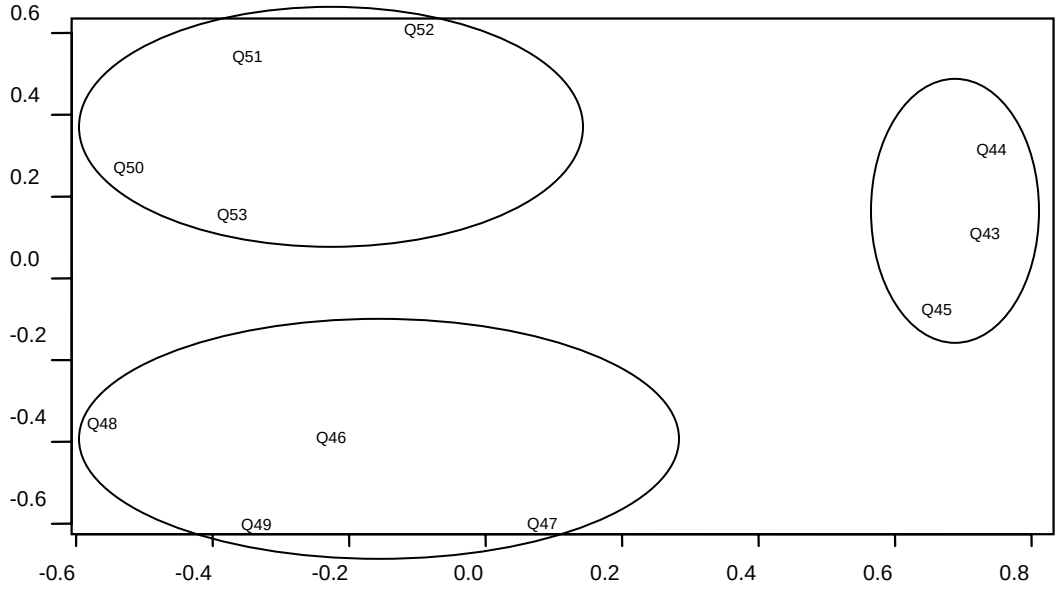
The HCA presented in Figure 2 shows that there are 3 larger clusters: Cluster 1: Q43, Q44 and Q45; Cluster #2: Q51, Q52, Q50 and Q53; Cluster #3: Q48, Q49, Q46, Q47. One problem with the HCA is that sometimes it forces clusters that are basically not very different

from each other. That is why additional step is necessary to further investigate the scalability before the final decision is made.

Step 2. Multidimensional Scaling (MDS)

Kruskal's non-metric multidimensional scaling is used for the analysis. This method is part of a large class of methods based on cases represented in a low-dimensional Euclidean space, usually 2-dimensional. The representation is such that the proximity of the points on the graph reflects the similarities of their variables. Multidimensional scaling will be used in close relation to the HCA described in Step 1 above.

Figure 3. Multidimensional Scaling



In this case the analysis of Figure 3 confirms the results from HCA that there are 3 distinctive clusters or subscales of this scale. In general the multidimensional scaling representation might show that some or all of the clustering is not necessary and thus the scale may be unidimensional.

D. Mokken Scale Analysis

We will perform Mokken scale analysis for polytomous and dichotomous scale items using nonparametric cumulative item response theory (IRT). We will test the scalability of a given scale or construct one or more unidimensional scales from an item pool. We will be able to present an evaluation of the model fit of a given scale, including the assessment of its reliability and suggestions for removal of misfitting items.

Example 1:

We are analyzing the same data as in the above examples. In this case we will use 0.4 as the threshold level for Loevinger's H for an item inclusion in the scale.

Scale 1. On the first round Scale 1 is formed and it contains all 11 items. No items are left for further analysis. The scale has 11 items with overall H=0.68 (min H=0.62) and reliability Rho=0.94. This is a strong scale with high reliability.

k	H	Rho
11	0.68	0.94

Item coefficients

Item	Label	Mean	ItemH
Q51		2.81	0.67
Q50		2.89	0.71
Q53		2.95	0.73
Q46		2.98	0.70
Q48		3.08	0.65
Q49		3.09	0.66
Q47		3.15	0.70
Q52		3.22	0.66
Q45		3.28	0.66
Q43		3.54	0.62
Q44		3.57	0.63

Example 2:

This time we are analyzing a scale with 18 items. We will use 0.3 as the threshold level for the Loevinger's H for an item's inclusion in the scale.

Scale 1a. On the first round, Scale 1 is formed and it contains 14 of the 18 items. Four items are excluded: LEQ1, LEQ14, LEQ16, LEQ17 due to the low level of the H coefficient. The scale has 14 items with overall H=0.37 (min H=0.3) and reliability Rho=0.87. This is a weak scale with high reliability.

Scale 1:

k	H	Rho
14	**0.37**	**0.87**

Item	Mean	ItemH
LEQ11	2.71	0.39
LEQ12	2.80	0.30
LEQ18	3.05	0.41
LEQ13	3.06	0.36
LEQ15	3.10	0.38
LEQ5	3.11	0.42
LEQ10	3.23	0.38
LEQ6	3.37	0.43
LEQ7	3.37	0.35
LEQ8	3.43	0.32
LEQ2	3.43	0.39
LEQ9	3.61	0.33
LEQ3	3.63	0.37
LEQ4	3.67	0.30
Excluded items:		
Item	**Mean**	**ItemH**
LEQ16	3.13	0.28
LEQ17	3.20	0.28
LEQ1	3.20	0.14
LEQ14	3.23	0.27

Scale 1b. On the next round of the search two of the leftover items comprise Scale 2: LEQ16 and LEQ17. This is a weak scale (H=0.31) with low reliability (Rho=0.48) and it will not be considered any further. The remaining two items LEQ1 and LEQ14 (with H=0.13 and H=0.25 respectively) are out of further consideration and do not belong to the scale.

Scale 2:

k	H	Rho
2	0.31	0.48

Item	Mean	ItemH
LEQ16	3.13	0.31
LEQ17	3.20	0.31
Excluded items:		
Item	**Mean**	**ItemH**
LEQ1	3.20	0.13
LEQ14	3.23	0.25

The conclusion is that only 14 out of the 18 items should be used to create a scale which is weak but very reliable.

E. Reliability and Scalability Relationship

The main combinations for a scale are presented in Table 4.

Table 5. Reliability and Scalability Dimensions

	Reliability	
Scalability	**Low** (Alpha < 0.7)	**High** (Alpha ≥ 0.7)
Weak (0.3≤ H < 0.4)	Bad scale	Measuring a construct that is not homogeneous
Moderate/Strong (H ≥ 0.4)	Too few items	Excellent scale: Measures one thing well.

Excluded from the above Table 5 are cases with Loevinger's H<0.3 because they cannot be considered scales in the Mokken sense.

Case 1: Weak scalability and low reliability. This is a case of a bad scale that should not be used in practice.

Case 2: Weak scalability and high reliability. This scale measures some phenomenon well, but it is not homogeneous enough. This scale can be improved by creating separate subscales or redefining the latent variable in question.

Case 3: Moderate to strong scalability and low reliability. This scale is technically very good, but the reliability could be improved significantly by including more questions/items.

Case 4: Moderate to strong scalability and high reliability. This is an excellent scale. It measures one phenomenon and it is well defined and homogeneous.

F. Final Reconciliation of the Scale Analysis

The three main components of the analysis - factor analysis, reliability and scalability - will guide us to make final decisions about each and every one of the scales.

In the example presented in the Methods section we can observe the following.

Table 6. Final Scale Analysis

	Method	Scale	Items	Loevinger's H	Reliability
Original Scale	All items included	Scale 0.	Q43-Q53	0.68	0.94
Suggested Scales	FA	Scale 1a.	Q43-Q45	0.84	0.87
		Scale 1b.	Q46-Q53	0.70	0.93
	HCA/MDS	Scale 2a	Q43-Q45	0.84	0.87
		Scale 2b	Q46-Q49	0.71	0.88
		Scale 2c	Q50-Q53	0.75	0.90
	Mokken Scale Analysis	Scale 3	Q43-Q53	0.68	0.94

Conclusion. The analyzed scale is strong and reliable. There is some evidence that groups of items might constitute subscales: Subscale 1: Items Q43-Q45; Subscale 2: Items Q46-Q49; and Subscale 3: Items Q50-Q53.

PART A.

ANALYSIS WITH COMBINED DATA FROM ALL SCHOOLS

A. Student Survey: 1. Overall Scale

The overall scale consists of all 43 items (LE 1-18, SPE 19-35, HSR 37-44).

Table 7. Student, Overall Scale Analysis

Method		LE Items (1-18)	SPE Items (19-35)	HSR Items (37-44)	H	Reliability
Factors (39.9% Var.)		**Factor Analysis**				
F1 (11.8%)	**LE**	5-8, 10-18	23, 27, 33-35		**0.31**	**0.87**
F2 (7.9%)	**HSR**			37-44	**0.40**	**0.80**
F3 (5.5%)	**SPE**		29-31		**0.60**	**0.81**
F4 (5.4%)	**SPE**		19-22		**0.50**	**0.78**
F5 (4.8%)	**LE**	2,3,4,9			**0.37**	**0.66**
F6 (4.5%)	**SPE**		24,25,32		**0.57**	**0.78**
Not Loaded*		1	26,28			
Double Loaded**		2 ,6, 12	22	39, 40		
		Reproducibility and Consistency				
Cronbach's Alpha		0.939				
Inconsistent Items		# 1				
Clusters		**Multidimensional Scaling**				
C1	**LE**	1-18	34		**0.31**	**0.87**
C2	**SPE**		19-33,35		**0.31**	**0.85**
C3	**HSR**			37-44	**0.40**	**0.80**
Scales		**Mokken Scale Analysis**				
All Items		1-18	19-35	37-44	**0.30**	**0.94**
Scale 1	**Mixed**	2,3,5-7,9-13,15,18	20,22,23, 26-31,33-35	37-40,42-44	**0.34**	**0.93**
Scale 2	**SPE**		19,21,24,25, 32		**0.43**	**0.77**
Not Included		1,4,8,14,16,17		41		
Modified Scale		2-18	19-35	37-44	**0.31**	**0.94**

* "Not Loaded" means that this item does not have any factor loadings greater than 0.3.
** "Double Loaded" means that this item has more than one loadings greater than 0.3. It is assigned to the factor with the largest loading.

The overall scale for students has **weak scalability and very high reliability**.

Item #1; "My classes are challenging (not too easy; they make me think)" should be excluded from the scale. It is inconsistent and stands out from the rest of the scale items.

The three original subscales LE, SPE, and HSR seem to hold mostly together and the SPE items may form 3 small subscales. The overall scale and most subscales are very reliable with Rho varying from 0.77 to 0.94. Only one subscale combining LE items 2-4, 9 is not very reliable. The strongest subscale (H=0.60, Rho=0.80) consists of items 29-31 which address safety concerns at school. Another strong subscale (H=0.50, Rho=0.78) combines items 19-22 which have to do with the school cleanliness.

Some SPE items tend to group with LE items, particularly item # 34 ("Teachers work together to help students at my school."

Another strong subscale (H=0.57, Rho=0.78) is also from SPE and it combines items 24, 25, 32 which have to do with students' behavior and how well they get along.

The items from HSR scale stay together almost always. This is the most stable and unidimensional scale that does not form subscales.

LE scale can be divided in 2 subscales but one is not reliable so we can consider the LE scale more or less unidimensional, weak and reliable. The SPE scale has a weak scalability as a whole but divides into 3 strong subscales. The HSR scale is unidimentional, moderately scalable and very reliable.

The overall scale is a scale in Mokken sense and it has weak scalability and high reliability. The scale will represent the whole school environment, including learning, social and physical environment and home-school relations.

Student Survey: 2. Learning Environment (LE) Scale

The LE scale consists of 18 items (LE 1-18).

Table 8. Student, LE Scale Analysis

Method		LE Items (1-18)	H	Reliability
Factors (34.6% Var.)		**Factor Analysis**		
F1	(13.6%)	10-18	**0.34**	**0.80**
F2	(11.1%)	5-7	**0.50**	**0.72**
F3	(9.9%)	2-4, 9	**0.42**	**0.71**
Not Loaded		1, 8		
Double Loaded		2, 6, 10, 11, 13, 18		
		Reproducibility and Consistency		
Cronbach's Alpha		0.870		
Inconsistent Items		# 1		
Clusters		**Multidimensional Scaling**		
C1		2-4, 9	**0.42**	**0.71**
C2		5-8	**0.42**	**0.71**
C3		10-18	**0.34**	**0.80**
C4		1		
Scales		**Mokken Scale Analysis**		
All Items		1-18	**0.31**	**0.87**
Scale 1		2-13, 15, 18	**0.37**	**0.87**
Not Included		1, 14, 16, 17		
Modified Scale		2-18	**0.34**	**0.88**

The LE scale for students has **weak scalability and very high reliability**. Item #1; "My classes are challenging (not too easy; they make me think)" is suggested for removal from the scale because it is inconsistent with the rest of the items and forms a separate cluster/subscale. After its removal the modified scale has slightly better scalability and reliability.

There are several possible subscales. The one with the strongest scalability (H=0.50) but with not very impressive reliability (Rho=0.72) consists of items 5-7 which have to do with teachers helping students and doing their job in teaching mathematics. The second subscale consists of items 10-18 and it has good reliability but the scalability is still weak.

Student Survey: 3. Social and Physical Environment (SPE) Scale

The SPE scale consists of 17 items (SPE 19-35).

Table 9. Student, SPE Scale Analysis

Method		SPE Items (19-35)	H	Reliability
Factors (46.5% Var.)		**Factor Analysis**		
F1	(14.7%)	22, 23, 26-28, 33-35	**0.38**	**0.81**
F2	(11.8%)	29-31	**0.66**	**0.85**
F3	(10.2%)	19-21	**0.56**	**0.77**
F4	(9.8%)	24, 25, 32	**0.52**	**0.74**
Not Loaded		None		
Double Loaded		22, 29, 32, 33		
		Reproducibility and Consistency		
Cronbach's Alpha		0.893		
Inconsistent Items		None		
Clusters		**Multidimensional Scaling**		
C1		24, 25, 32, 33	**0.48**	**0.77**
C2		19-23	**0.45**	**0.77**
C3		29-31	**0.66**	**0.85**
C4		26-28, 34, 35	**0.38**	**0.74**
Scales		**Mokken Scale Analysis**		
All Items		19-35	**0.37**	**0.90**
Scale 1		19-35	**0.37**	**0.90**
Not Included		none		

The SPE scale for students has **weak to moderate scalability and very high reliability**. It is clearly a multidimensional scale. The best subscale includes items 29-31 that deal with safety issues at school. This subscale although very small is very strong and very reliable. Another subscale consists of items 19-21 that are concerned with the school being kept clean. The other two possible subscales have weak to moderate scalability and very good reliability. No modifications are proposed for the SPE scale since there are no items with obvious inconsistency.

Student Survey: 4. Home-School Relations (HSR) Scale

The HSR scale consists of 8 items (HSR 37-44).

Table 10. Student, HSR Scale Analysis

Method		HSR Items (37-44)	H	Reliability
Factors (36.5% Var.)		**Factor Analysis**		
F1	(36.5%)	37-44	**0.41**	**0.82**
Not Loaded		None		
Double Loaded		None		
		Reproducibility and Consistency		
Cronbach's Alpha		0.816		
Inconsistent Items		None		
Clusters		**Multidimensional Scaling**		
C1		37, 38, 41	**0.45**	**0.67**
C2		39,40	**0.49**	**0.62**
C3		42-44	**0.42**	**0.63**
Scales		**Mokken Scale Analysis**		
All Items		37-44	**0.41**	**0.82**
Scale 1		37-44	**0.41**	**0.82**
Not Included		None		

The HSR scale for students has **moderate scalability and high reliability**. This is definitely a unidimensional scale. There are 3 possible subscales suggested by the MDS but they are not confirmed by the other methods and the subscales have reliability below the threshold of 0.7. No modifications are necessary for this scale.

B. Teacher Survey: 1. Overall Scale

The overall scale consists of all 53 items (LE 1-26, SPE 27-42, HSR 43-53).

Table 11. Teacher, Overall Scale Analysis

Method		LE Items (1-26)	SPE Items (27-42)	HSR Items (43-53)	H	Reliability
Factors (54.3% Var.)		**Factor Analysis**				
F1 (16.1%)	**HSR**	13	32,33,39,40,42	43-53	**0.62**	**0.95**
F2 (14.3%)	**LE**	10-12, 14, 17-26	34, 35, 41		**0.55**	**0.94**
F3 (10.7%)	**LE**	1-9			**0.52**	**0.89**
F4 (7.1%)	**SPE**		27-31		**0.63**	**0.87**
F5 (6.0%)	**SPE**		36-38		**0.46**	**0.97**
Not Loaded		15, 16				
Double Loaded		7, 10, 11, 13, 21, 23, 24, 26,	34, 35, 39	40, 41, 42, 43, 44, 45		
		Reproducibility and Consistency				
Cronbach's Alpha		0.970				
Inconsistent Items						
Clusters		**Multidimensional Scaling**				
C1	**LE**	1-12, 14-26			**0.47**	**0.95**
C2	**SPE**		32-42		**0.62**	**0.92**
C3	**SPE**		27-31		**0.63**	**0.87**
C4	**HSR**	13		43-53	**0.67**	**0.94**
Scales		**Mokken Scale Analysis**				
All Items		1-26	27-42	43-53	**0.46**	**0.97**
Scale 1	**Mixed**	1-4, 6,7, 9-14, 17-26	32-42	43-53	**0.51**	**0.97**
Scale 2	**SPE**		27-31		**0.63**	**0.87**
Not Included		5,8,15,16				
Modified Scale		1-14, 17-26	27-30, 32-42	43-53	**0.48**	**0.97**

The overall scale for teachers has **moderate to strong scalability and very high reliability**. Items 15 and 16 are suggested for removal because they are inconsistent with the rest of the scale.

LEQ15	Our school has a good selection of library and media material.
LEQ16	Our school has sufficient computers for instructional use.

They seem to be important questions related to computer use and library but they do not seem to belong to this scale. After their removal the scalability rises slightly and the reliability stays the same.

The LE scale is strong and reliable and although there is a possibility for 2 subscales they do not stand out according to the MDS analysis. Item 13 seems not to be part of the LE scale – it loads with HSR items.

LEQ13	Students at my school are motivated and interested in learning.

There is no obvious explanation for this phenomenon.

The SPE scale seems to have more than one dimension. Two of the possible 3 subscales have strong scalability and very high reliability. One subscale combines items 27-31 that have to do with the school being clean. Another strong subscale includes items 32-42 that have to do with safety, good behavior, and getting along well.

The HSR scale is very strong and very reliable unidimensional scale. There are no subscales or dimensions suggested by any of the methods used.

Teacher Survey: 2. Learning Environment (LE) Scale

The LE scale consists of 26 items (LE 1-26).

Table 12. Teacher, LE Scale Analysis

Method		LE Items (1-26)	H	Reliability
Factors (52.0% Var.)		**Factor Analysis**		
F1	(23.7%)	10-13, 18-26	**0.60**	**0.94**
F2	(18.4%)	1-9	**0.52**	**0.89**
F3	(9.9%)	14-17	**0.52**	**0.80**
Not Loaded		None		
Double Loaded		7, 11, 13, 14, 18, 20, 21, 23, 24, 26,		
		Reproducibility and Consistency		
Cronbach's Alpha		0.944		
Inconsistent Items		# 15, 16		
Clusters		**Multidimensional Scaling**		
C1		1-9	**0.52**	**0.89**
C2		14-18	**0.48**	**0.81**
C3		10-13, 19-26	**0.62**	**0.94**
Scales		**Mokken Scale Analysis**		
All Items		1-26	**0.47**	**0.95**
Scale 1		1-4, 6-14, 17-26	**0.51**	**0.95**
Not Included		5, 15, 16		
Modified Scale		1-14, 17-26	**0.50**	**0.95**

The LE scale for teachers has **moderate scalability and very high reliability**. Items 15 and 16 are suggested for exclusion because they are inconsistent with the rest of the items although they are very interesting questions. Without the 2 items in question the modified scale has stronger scalability while the reliability stays the same.

There are a couple of possible subscales. One subscale combines items 1-9 that have to do with effective instructional strategies. Another possible strong subscale combines items10-13 and 18-26 that have to do with teachers respect of each other and school administration and teacher evaluation. The Mokken scale analysis though points to a predominantly single scale.

Teacher Survey: 3. Social and Physical Environment (SPE) Scale

The SPE scale consists of 16 items (SPE 27-42).

Table 13. Teacher, SPE Scale Analysis

Method		SPE Items (27-42)	H	Reliability
Factors (60.7% Var.)		**Factor Analysis**		
F1	(23.9%)	32-35, 39-42	**0.62**	**0.91**
F2	(20.5%)	27-31	**0.63**	**0.87**
F3	(16.3%)	36-38	**0.85**	**0.90**
Not Loaded		None		
Double Loaded		36, 37, 39, 40, 42		
		Reproducibility and Consistency		
Cronbach's Alpha		0.921		
Inconsistent Items		# 31		
Clusters		**Multidimensional Scaling**		
C1		27-31	**0.63**	**0.87**
C2		36-38	**0.85**	**0.90**
C3		39-42	**0.58**	**0.83**
C4		32-35	**0.73**	**0.88**
Scales		**Mokken Scale Analysis**		
All Items		27-42	**0.51**	**0.93**
Scale 1		27-30, 32-42	**0.53**	**0.93**
Not Included		31 (H=0.39)		
Modified Scale		27-30, 32-42	**0.53**	**0.93**

The SPE scale for teachers has **strong scalability and very high reliability**. Only item # 31 is suggested for removal because it is to a certain degree inconsistent with the rest of the items.

SPEQ31	There is sufficient space for instructional programs at my school.

After the removal of this item the modified scale has better scalability and the reliability remains very high.

One of the possible small subscales has extremely high scalability and reliability. It combines items 36-38 that have to do with school safety. Another possible strong and reliable subscale includes items 27-31 that deal with a clean and well maintained school.

The Mokken scale analysis though suggests one scale only with item # 31 left out.

Teacher Survey: 4. Home-School Relations (HSR) Scale

The HSR scale consists of 11 items (HSR 43-53).

Table 14. Teacher, HSR Scale Analysis

Method		HSR Items (43-53)	H	Reliability
Factors (66.9% Var.)		**Factor Analysis**		
F1	(43.2%)	46-53	**0.72**	**0.94**
F2	(23.7%)	43-45	**0.83**	**0.89**
Not Loaded		None		
Double Loaded		45, 46, 47, 52, 53		
		Reproducibility and Consistency		
Cronbach's Alpha		0.937		
Inconsistent Items		None		
Clusters		**Multidimensional Scaling**		
C1		43-45	**0.83**	**0.89**
C2		46-49	**0.71**	**0.89**
C3		50-53	**0.77**	**0.91**
Scales		**Mokken Scale Analysis**		
All Items		43-53	**0.68**	**0.94**
Scale 1		43-53	**0.68**	**0.94**
Not Included		None		

The HSR scale for teachers has **very strong scalability and very high reliability**. None of the items is inconsistent and no modifications are suggested for this scale. There are a few possible small subscales but they do not stand out with very different characteristics than the overall scale. One possible subscale combines items 43-45 that deal with parents' knowledge of the school programs and policies. Another possible subscale includes items 46-49 that deal with parents' cooperation and interest in their children schoolwork.

No modifications are suggested for this scale.

C. Parent Survey: 1. Overall Scale

The overall scale consists of all 46 items.

The True/False subscale is not suitable for FA so it is excluded from it.

Table 15. Parent, Overall Scale Analysis

Method		LE 1-5	SPE 1-5	HSR 1-11	Sec4 1-8	Sec 5 1-5	Rate 1-5	TF 1-7	H	Reliability
Factors (52.7% Var.)		**Factors Analysis**								
F1 (15.4%)	**HSR**		3	1-11					**0.65**	**0.94**
F2 (8.8%)	**Rate**						1-5		**0.83**	**0.95**
F3 (8.2%)	**SPE**		1,2,4, 5						**0.75**	**0.88**
F4 (7.4%)	**Sec 4**				3-8	1			**0.46**	**0.83**
F5 (7.0%)	**LE**	1-5							**0.70**	**0.90**
F6 (5.9%)	**Sec 5**				1,2	2-5			**0.46**	**0.79**
Not Loaded										
Double Loaded		3,4,5	3,4,5	9,10, 11			1-5			
		Reproducibility and Consistency								
Cronbach's Alpha		0.769								
Inconsistent Items										
Cluster		**Multidimensional Scaling**								
C1	**Mixed**	1-5	1-5	1-11			1-5		**0.20**	**0.87**
C2	**Sec 4&5**				1-8	1-5			**0.43**	**0.86**
No Cluster								1-7	**0.26**	**0.58**
Scales		**Mokken Scale Analysis**								
All Items		1-5	1-5	1-11	1-8	1-5	1-5	1-7	**0.10**	**0.82**
Scale 1	**Rate**						1-5		**0.83**	**0.95**
Scale 2	**Mixed**	1-5	1-5	1-11				5-7	**0.59**	**0.96**
Scale 3	**Sec 4&5**				1-7	1-5			**0.44**	**0.85**
Not Included					8 H=0.4			1-4 H<0		
Proposed new scales										
School Environment Scale		1-5	1-5	1-11					**0.61**	**0.96**
Sections 4 and 5 Scale					1-8	1-5			**0.43**	**0.86**
Rate Scale							1-5		**0.83**	**0.95**
Reduced True/False Scale								5-7	**0.61**	**0.71**

The full overall scale for parents including all 46 items **is not a scale** in Mokken sense.

The analysis shows that the overall scale should be divided in 4 separate scales.

1. School Environment scale that includes LE, SPE and HSR scales
2. Sections 4 and 5 scale
3. Rate scale
4. True/False section scale

Without any doubt the True/False scale is the worst of all scales not only for parent but also for student and teacher surveys. For now we can say that items 1-4 of the True/False scale have to be removed. If necessary items 5-7 may remain.

The modified School Environment scale has **strong scalability and very high reliability**. The combined Section 4 and 5 scale has **moderate scalability and high reliability**. The Rate scale is the best of all with **very high scalability and high reliability**. The reduced True/False scale has strong scalability but barely enough reliability and can not be recommended for use in practice. Later on we will investigate it further.

Parent Survey: 2. Learning Environment (LE) Scale

The LE scale consists of 5 items (LE 1-5).

Table 16. Parent, LE Scale Analysis

Method		LE Items (1-5)	H	Reliability
Factors (58.2% Var.)		**Factor Analysis**		
F1	(58.2%)	1-5	**0.66**	**0.88**
Not Loaded		None		
Double Loaded		None		
		Reproducibility and Consistency		
Cronbach's Alpha		0.876		
Inconsistent Items		None		
Clusters		**Multidimensional Scaling**		
C1		1		
C2		2, 5	**0.74**	**0.79**
C3		3,4	**0.76**	**0.79**
Scales		**Mokken Scale Analysis**		
All Items		1-5	**0.66**	**0.88**
Scale 1		1-5	**0.66**	**0.88**
Not Included		None		

The LE scale for parents has **strong scalability and high reliability**. There are no inconsistent items and this scale should be kept intact. No modifications are suggested.

Parent Survey: 3. Social and Physical Environment (SPE) Scale

The SPE scale consists of 5 items (SPE 1-5).

Table 17. Parent, SPE Scale Analysis

Method		SPE Items (1-5)	H	Reliability
Factors (55.7% Var.)		**Factor Analysis**		
F1	(55.7%)	1-5	**0.66**	**0.88**
Not Loaded		None		
Double Loaded		None		
		Reproducibility and Consistency		
Cronbach's Alpha		0.866		
Inconsistent Items		None		
Clusters		**Multidimensional Scaling**		
C1		1,2	**0.66**	**0.77**
C2		3		
C3		4,5	**0.81**	**0.82**
Scales		**Mokken Scale Analysis**		
All Items		1-5	**0.66**	**0.88**
Scale 1		1-5	**0.66**	**0.88**
Not Included		None		

The SPE scale for parents has **strong scalability and high reliability**. There are no inconsistent items and this scale should be kept intact. No modifications are suggested.

Parent Survey: 4. Home School Relations (HSR) Scale

The HSR scale consists of 11 items (HSR 1-11).

Table 18. Parent, HSR Scale Analysis

Method		HSR Items (1-11)	H	Reliability
Factors (51.7% Var.)		**Factor Analysis**		
F1	(51.7%)	1-11	**0.63**	**0.94**
Not Loaded		None		
Double Loaded		None		
		Reproducibility and Consistency		
Cronbach's Alpha		0.934		
Inconsistent Items		None		
Clusters		**Multidimensional Scaling**		
C1		1-3	**0.73**	**0.87**
C2		5-7	**0.74**	**0.86**
C3		4, 8-11	**0.61**	**0.87**
Scales		**Mokken Scale Analysis**		
All Items		1-11	**0.63**	**0.94**
Scale 1		1-11	**0.63**	**0.94**
Not Included		None		

The HSR scale for parents has **strong scalability and very high reliability**. There are no inconsistent items. There are several possible subscales suggested by MDS but they are very small scales and they do not present enough differences with the overall scale.

No modifications are suggested.

Parent Survey: 5. Section 4 Scale

The Section 4 scale consists of 8 items (Sec4 1-8).

Table 19. Parent, Section 4 Scale Analysis

Method		Section 4 Items (1-8)	H	Reliability
Factors (45.0% Var.)		**Factor Analysis**		
F1	(27.1%)	3-8	**0.49**	**0.82**
F2	(17.9%)	1, 2	**0.53**	**0.68**
Not Loaded		None		
Double Loaded		3, 4, 6		
		Reproducibility and Consistency		
Cronbach's Alpha		0.814		
Inconsistent Items		None		
Clusters		**Multidimensional Scaling**		
C1		1,2	**0.53**	**0.68**
C2		3,4	**0.55**	**0.68**
C3		5-8	**0.55**	**0.78**
Scales		**Mokken Scale Analysis**		
All Items		1-8	**0.45**	**0.82**
Scale 1		3-8	**0.49**	**0.82**
Scale 2		1,2	**0.53**	**0.68**
Not Included		None		

The Section 4 scale for parents has **moderate scalability and very good reliability**. There are two small possible subscales. One combines two items 1 and 2 that deal with parents attending parent-teacher conferences and student programs. Another possible subscale includes items 3-8 that deal with a variety of issues including volunteering and attending workshops. No modifications are proposed for this scale.

Parent Survey: 6. Section 5 Scale

The Section 5 scale consists of 5 items (Sec5 1-5).

Table 20. Parent, Section 5 Scale Analysis

Method		Section 5 Items (1-5)	H	Reliability
Factors (34.5% Var.)		**Factor Analysis**		
F1	(34.5%)	1-5	**0.47**	**0.71**
Not Loaded		None		
Double Loaded		None		
		Reproducibility and Consistency		
Cronbach's Alpha		0.658		
Inconsistent Items		# 1		
Clusters		**Multidimensional Scaling**		
C1		1,2	**0.52**	**0.59**
C2		3-5	**0.55**	**0.70**
Scales		**Mokken Scale Analysis**		
All Items		1-5	**0.47**	**0.71**
Scale 1		1-5	**0.47**	**0.71**
Not Included		None		

This Section 5 scale for parents has **moderate scalability and barely acceptable reliability**. The scalability might improve slightly if items 1 and 2 are removed but then the scale would become very small.

Parent Survey: 7. Rate Scale

The Rate scale consists of 5 items (Rate 1-5).

Table 21. Parent, Rate Scale Analysis

Method		Section 5 Items (1-5)	H	Reliability
Factors (71.5% Var.)		**Factor Analysis**		
F1	(71.5%)	1-5	**0.78**	**0.94**
Not Loaded		None		
Double Loaded		None		
		Reproducibility and Consistency		
Cronbach's Alpha		0.927		
Inconsistent Items		None		
Clusters		**Multidimensional Scaling**		
C1		1,5	**0.78**	**0.85**
C2		3,4	**0.86**	**0.90**
C3		2		
Scales		**Mokken Scale Analysis**		
All Items		1-5	**0.78**	**0.94**
Scale 1		1-5	**0.78**	**0.94**
Not Included		None		

The Rate scale for parents has **very strong scalability and very high reliability**. This is one of the best scales from the 3 surveys. The scale should be kept intact and no modifications are suggested.

Parent Survey: 8. True/False Scale

The True/False scale consists of 7 items (TF 1-5).

Dichotomous items can not be analyzed by FA so it is excluded from the analysis.

Table 22. Parent, True/False Scale Analysis

Method	TF Items (1-7)	H	Reliability
	Reproducibility and Consistency		
Cronbach's Alpha	0.543		
Inconsistent Items	# 4		
Clusters	**Multidimensional Scaling**		
C1	1-3	**0.30**	**0.54**
C2	5-7	**0.61**	**0.74**
C3	4		
Scales	**Mokken Scale Analysis**		
All Items	1-7	**0.22**	**0.57**
Scale 1	5-7	**0.61**	**0.74**
Not Included	1-4 (H=0.11-0.15)		

The TF scale for parent **is not a scale** in Mokken sense. A reduced scale that includes only items 5-7 has strong scalability and modest reliability but it has only 3 items. Our suggestion is to drop the scale from the survey. If there is a need, modifications can be done both changing the questions and possible answers, most probably moving from Yes/No to a Likert-type scale.

PART B.

ANALYSIS BY SCHOOL TYPE

A. Student Survey

The focus of this analysis will be to compare the scale results by school type and also to compare them to the results using the combined data from all schools.

Table 23. Student Overall Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
FA	Factors	6	6	6	7	7	7
	Variance	39.9%	34.2%	36.9%	40.7%	44.7%	42.3%
RC	Alpha	0.939	0.926	0.931	0.938	0.943	0.910
	Incons.	# 1	# 1	# 1	# 1	# 1	# 1
MDS	Clusters	3	3-4	Unclear	3	3	Unclear
MO H/Rho	Scale 1	0.34/0.93	0.34/0.90	0.34/0.91	0.34/0.93	0.38/0.95	0.36/0.88
	Scale 2	0.43/0.77	0.31/0.77	0.49/0.71	0.42/0.79	0.37/0.69	0.45/0.78
	Scale 3			0.36/0.77			0.35/0.85
	Original Scale	0.30/0.90	0.27/0.93	0.27/0.93	0.29/0.94	0.32/0.95	0.24/0.92
	Modified* Scale	0.31/0.94	0.28/0.93	0.28/0.93	0.29/0.94	0.32/0.95	0.24/0.92

* Item # 1 is excluded.

The results in **Table 23** show that overall there are few differences by school type and in comparison to the combined results. The factor analysis produces very similar results and the explained variance for all of the models seems to be not very high, around 40%. Item # 1 is suggested for removal from all of them.

The hierarchical cluster analysis results and multidimensional scaling are not very clear for the high schools and the special schools. Not surprisingly the Mokken scale analysis brought up 3 scales for the same two types compared to 2 scales for the rest of the schools.

The modified scales have weak scalability and very high reliability. In the strict sense, regarding the threshold we have established (H=0.3), the scalability for Elementary, Middle, High, and Special schools is very low. These scales can be used with caution.

Table 24 shows that for the LE scale there are hardly any differences between the different school types and combined data results. Four of the tests suggest removal of item #1, while for the career centers the inconsistent item is # 16:

LEQ16	The media center at my school has a good selection of books.

This is probably something specific to the career centers where media centers my not be readily available so this question might not be relevant.

Table 24. Student LE Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
FA	Factors	3	3	3	3	3	3
	Variance	34.6%	29.5%	31.6%	35.1%	39.1	34.6
RC	Alpha	0.870	0.829	0.849	0.874	0.873	0.859
	Incons.	# 1	#1	#1	#1	#16	#4
MDS	Clusters	4	4	3	3	3	3
MO H/Rho	Scale 1	0.37/0.87	0.33/0.81	0.37/0.83	0.37/0.87	0.38/0.87	0.35/0.85
	Original Scale	0.31/0.87	0.25/0.84	0.27/0.85	0.32/0.88	0.31/0.88	0.29/0.86
	Modified* Scale	0.34/0.88	0.27/0.84	0.29/0.85	0.33/0.88	0.32/0.88	0.31/0.86

* Item # 1 is excluded.

For the special schools item #4 is inconsistent:

LEQ4	My teachers expect students to behave.

Again, it might be the case that students in special schools are under different rules and regulations and this question might be not quite relevant for them.

The modified scales have weak scalability and high reliability. Technically two of the scales (for Elementary and Middle schools) are below the threshold of H=0.3. These scales should be used with caution.

Table 25. Student SPE Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
FA	Factors	4	3	4	4	4	4
	Variance	36.5%	39.2%	46.5%	47.8%	52.2%	46.4%
RC	Alpha	0.893	0.880	0.889	0.901	0.912	0.854
	Incons.					#35	
MDS	Clusters	4	5	3	4	3	4
MO H/Rho	Scale 1	0.37/0.90	0.36/0.88	0.36/0.89	0.38/0.90	0.44/0.92	0.35/0.87
	Original Scale	0.37/0.90	0.35/0.89	0.36/0.89	0.38/0.90	0.41/0.91	0.28/0.86

The factor analysis results in **Table 25** show that the explained variance varies from 39% for the elementary schools to 52% for the career centers. It might be the case that for students in career centers the social and physical environment is of greater importance than in elementary schools.

Also for the career centers item # 35 is inconsistent with the rest of the items:

SPEQ35	I am satisfied with the social and physical environment at my school.

This is the last question of this scale and is supposed to capture the overall opinion of the student regarding the social and physical environment. This seems to be difficult for the students in career centers whose answers to the previous questions from this scale are not consistent with the last one.

Most scales have weak scalability with the exception of the career centers with moderate strength scalability. The reliability is very high for all scales.

The factor analysis in **Table 26** shows explained variance from 30% for the elementary schools to 49% for the special schools. All scales have good reliability and the scalability varies from weak to moderate.

Table 26. Student HSR Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
FA	Factors	1	1	1	1	1	3
	Variance	36.5%	30.2%	32.7%	36.5%	38.3	49.2
RC	Alpha	0.816	0.762	0.787	0.815	0.826	0.781
	Incons.						
MDS	Clusters	3	2	3	3	2	3
MO H/Rho	Scale 1	0.41/0.82	0.34/0.75	0.36/0.79	0.41/0.82	0.42/0.83	0.37/0.79
	Original Scale	0.41/0.82	0.33/0.76	0.36/0.79	0.41/0.82	0.42/0.83	0.37/0.79

B. Teacher Survey

Table 27. Teacher Overall Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Primary Schools	Special Schools
FA	Factors	5	5	5	5	5	5	4
	Variance	54.3%	52.6%	53.5%	54.0%	53.8%	55.3%	48.3%
RC	Alpha	0.97	0.967	0.969	0.969	0.965	0.971	0.949
	Incons.							
MDS	Clusters	4	3	Unclear	5	Unclear	Unclear	Unclear
MO H/Rho	Scale 1	0.51/0.97	0.51/0.97	0.49/0.97	0.50/0.97	0.49/0.97	0.52/0.97	0.46/0.95
	Scale 2	0.63/0.87	0.61/0.85	0.63/0.87	0.65/0.87	0.62/0.87	0.45/0.68	0.53/0.87
	Scale 3		0.55/0.76		0.58/0.78	0.58/0.77		0.49/0.85
	Scale 4		0.49/0.78					
	Original Scale	0.46/0.97	0.44/0.97	0.44/0.97	0.45/0.97	0.43/0.97	0.49/0.97	0.32/0.95
	Modified* Scale	0.48/0.97	0.46/0.97	0.46/0.97	0.46/0.97	0.45/0.97	0.52/0.97	0.34/0.95

* Items # 15, 16 and 31 are excluded.

Factor analysis results in **Table 27** are similar with explained variance around 50%. Compared to the overall student scales, these are much better models.

The overall reliability for all scales is very high. The scalability varies from moderate to very strong. One interesting phenomenon is the fact that there were no clear patters from the multidimensional scaling for some of the school types. The Mokken scale analysis suggests a variety of scales from 2 to 4 possible different scales. The modified scales have strong scalability and very high reliability. Only the special schools scale has weak scalability.

Factor analysis results in **Table 28** show no particular differences by type of school and the overall reliability is very high for all scales. For most scales items # 15 and 16 seem to be inconsistent with the rest of the scale.

LEQ15	Our school has a good selection of library and media material.
LEQ16	Our school has sufficient computers for instructional use.

As mentioned before, these two questions are very important but they don't seem to belong to this scale.

Table 28. Teacher LE Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Primary Schools	Special Schools
FA	Factors	3	4	3	3	4	3	4
	Variance	52.0%	53.5%	51.1%	51.8%	55.0%	55.0%	48.7%
RC	Alpha	0.944	0.939	0.943	0.945	0.942	0.947	0.918
	Inconsist.	#15,16	#15,16	#16	#16	#5,15,16	#16	#5,8,15, 16
MDS	Clusters	3	3	3	2	3	Unclear	Unclear
MO H/Rho	Scale 1	0.51/0.95	0.50/0.94	0.50/0.95	0.50/0.95	0.51/0.95	0.53/0.95	0.49/0.93
	Scale 2					0.58/0.77		0.52/0.74
	Original Scale	0.47/0.95	0.45/0.94	0.46/0.95	0.47/0.95	0.46/0.95	0.50/0.95	0.36/0.92
	Modified* Scale	0.50/0.95	0.48/0.94	0.49/0.95	0.50/0.95	0.49/0.95	0.53/0.95	0.39/0.92

* Items # 15 and 16 are excluded.

For the career centers and special schools item # 5 is also inconsistent with the rest of the items.

LEQ5	There is a sufficient amount of classroom time allocated to instruction in essential skills.

Special schools have one more inconsistent item, #8.

LEQ8	My school offers effective programs for students with disabilities.

This item is obviously very important particularly for the special schools and that is why it does not fit well with the scale.

The modified scales have very strong scalability and very high reliability except the special schools' scale which has weak scalability.

Factor analysis's explained variance in **Table 29** is about 60% with 3 factors for most scales except 4 for the special schools. Overall reliability is very high with moderate to strong scalability.

Item # 31 is inconsistent across the board except for the special schools.

SPEQ31	There is sufficient space for instructional programs at my school.

SPE scale for special schools seems to be different from the rest of the school types.

The modified scales have moderate to strong scalability and very high reliability.

Table 29. Teacher SPE Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Primary Schools	Special Schools
FA	Factors	3	3	3	3	3	3	4
	Variance	60.7%	59.3%	60.1%	60.3%	56.0%	60.5%	63.2%
RC	Alpha	0.921	0.914	0.920	0.920	0.897	0.922	0.882
	Inconsist.	#31	#31	#31	#31	#31	#31	
MDS	Clusters	4	4	4	4	3	3	3
MO H/Rho	Scale 1	0.53/0.93	0.51/0.93	0.52/0.92	0.52/0.92	0.48/0.91	0.54/0.93	0.55/0.88
	Scale 2							0.57/0.82
	Scale 3							0.60/0.72
	Original Scale	0.51/0.93	0.49/0.92	0.50/0.92	0.50/0.52	0.45/0.91	0.54/0.93	0.40/0.89
	Modified* Scale	0.53/0.93	0.51/0.93	0.52/0.92	0.52/0.92	0.48/0.91	0.56/0.93	0.42/0.89

* Item # 31 is excluded.

The results from the factor analysis in **Table 30** show explained variance of 60% or more which is a very good indicator. Only 2 factors are extracted. The overall reliability and the scalability are very high. Item # 44 is inconsistent for the primary and special schools.

HSRQ44	Parents at my school know about school activities.

Table 30. Teacher HSR Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Primary Schools	Special Schools
FA	Factors	2	2	2	2	2	2	2
	Variance	66.9%	66.2%	65.8%	65.0%	61.7%	64.6%	61.0%
RC	Alpha	0.937	0.934	0.933	0.934	0.923	0.933	0.914
	Incons.						#44	#44
MDS	Clusters	3	3	3	3	3	3	4
MO H/Rho	Scale 1	0.68/0.94	0.67/0.94	0.65/0.94	0.65/0.93	0.61/0.93	0.67/0.94	0.59/0.92
	Original Scale	0.68/0.94	0.67/0.94	0.65/0.94	0.65/0.93	0.61/0.93	0.67/0.94	0.59/0.92

C. Parent Survey

Table 31. Parent Overall Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
FA	Factors	6	7	7	6	6	6
	Variance	52.7	52.6%	51.7%	52.2%	54.7%	57.6%
RC	Alpha	0.769	0.792	0.742	0.771	0.817	0.787
	Incons.						
MDS	Clusters	2	2	2	2	2	Unclear
MO H/Rho	Scale 1	0.83/0.95	0.83/0.95	0.81/0.94	0.80/0.94	0.58/0.96	0.85/0.95
	Scale 2	0.59/0.96	0.59/0.96	0.55/0.95	0.56/0.96	0.78/0.93	1.00/0.68
	Scale 3	0.44/0.85	0.50/0.79	0.50/0.69	0.47/0.89	0.52/0.91	0.63/0.96
	Scale 4			0.48/0.82			0.54/0.89
	Modified Scales						
	School Env.*	0.61/0.96	0.61/0.96	0.56/0.95	0.57/0.96	0.59/0.96	0.63/0.96
	Sections 4 & 5	0.43/0.86	0.38/0.82	0.40/0.85	0.46/0.89	0.52/0.91	0.48/0.88
	Rate	0.83/0.95	0.83/0.95	0.81/0.94	0.80/0.94	0.78/0.93	0.85/0.95
	Reduced TF**	0.61/0.71	0.55/0.69	0.65/0.72	0.64/0.72	0.58/0.69	0.55/0.66

* Includes items LE 1-5, SPE 1-5, HSR 1-11.

** Includes items TF 5-7

The factor analysis in **Table 31** suggests 6 or 7 factors that explain about 52% of the total variance of the models. The overall reliability is above the 0.7 threshold with highest reliability for the career centers. The Mokken scale analysis shows that there are subscales of the overall scale. As we know from the previous analysis of the combined data this result is very intuitive.

The first of the modified scales includes all 3 scales related to school environment. The scale has very high scalability for all types of schools and very high reliability. The second modified scale combines Sections 4 and 5. The scale has moderate to strong scalability with the highest value for the career centers. Reliability is uniformly high for all schools types. The Rate scale again is the best scale of all. It has scalability of 0.8 or more and reliability above 0.9. The True/False scale is again the worst of all. When it is reduced to 3 items (5-7) the scale has strong scalability and reliability around the threshold value of 0.7.

Table 32. Parent LE Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
FA	Factors	1	1	1	1	1	1
	Variance	58.2%	59.1%	56.0%	54.7%	55.4%	67.3%
RC	Alpha	0.876	0.879	0.867	0.863	0.868	0.917
	Incons.						
MDS	Clusters	3	3	3	3	2	3
MO H/Rho	Scale 1	0.66/0.88	0.67/0.89	0.64/0.87	0.62/0.87	0.63/0.88	0.81/0.92
	Original Scale	0.66/0.88	0.67/0.89	0.64/0.87	0.62/0.87	0.63/0.88	0.81/0.92

Table 32 shows pretty much uniform results. The factor analysis implies only one factor with 55% to 67% explained variance. The highest number is for the special schools scale which also scores the highest reliability (0.9) and scalability (0.8). The rest of the scales are also good and the scalability is strong, about 0.6.

Table 33. Parent SPE Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
FA	Factors	1	1	1	1	1	1
	Variance	55.7%	54.8%	52.4%	51.7%	57.1%	61.6%
RC	Alpha	0.866	0.859	0.849	0.850	0.877	0.883
	Incons.						
MDS	Clusters	3	3	3	3	3	3
MO H/Rho	Scale 1	0.66/0.88	0.67/0.88	0.62/0.86	0.61/0.86	0.68/0.89	0.70/0.88
	Original Scale	0.66/0.88	0.67/0.88	0.62/0.86	0.61/0.86	0.68/0.89	0.70/0.88

The SPE scale results in **Table 33** are very similar to the LE scale. Special schools scale stands out although the rest of the scales are also good. The scalability is very strong with high reliability. No inconsistent items are present.

Table 34. Parent HSR Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
FA	Factors	1	1	1	2	1	2
	Variance	51.7%	51.8%	48.0%	54.2%	50.0%	58.6%
RC	Alpha	0.934	0.934	0.925	0.926	0.930	0.930
	Incons.						#8
MDS	Clusters	3	3	3	3	4	3
MO H/Rho	Scale 1	0.63/0.94	0.63/0.93	0.60/0.93	0.62/0.93	0.62/0.93	0.63/0.94
	Original Scale	0.63/0.94	0.63/0.93	0.60/0.93	0.62/0.93	0.62/0.93	0.63/0.94

In **Table 34** the explained variance from the factor analysis is over 50% with only 1 or 2 factors extracted. The reliability is very high with high scalability. For the special schools item # 8 seems to be inconsistent.

HSRQ8	My child's school schedules activities at times that I can attend.

Table 35. Parent Section 4 Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
FA	Factors	2	2	2	2	2	2
	Variance	45.0%	41.2%	43.4%	50.5%	56.2%	52.8%
RC	Alpha	0.814	0.778	0.801	0.860	0.874	0.848
	Incons.					#1	#1,2
MDS	Clusters	3	3	3	Unclear	Unclear	2
MO H/Rho	Scale 1	0.49/0.82	0.49/0.78	0.47/0.81	0.52/0.88	0.56/0.88	0.59/0.87
	Scale 2	0.53/0.68		0.53/0.70			
	Original Scale	0.45/0.82	0.41/0.80	0.43/0.81	0.52/0.88	0.56/0.88	0.52/0.85

Section 4 scale factor analysis in **Table 35** reveals 2 factors for all scales with explained variance between 45% and 55%. Reliability is high for all scales. Multidimensional scaling does not reveal clear patterns for high schools and career centers. Scalability is from moderate to strong with high reliability.

Item # 1 is inconsistent with the rest of the scale for the career centers and special schools and item # 2 is also inconsistent for the special schools scale.

Sec4Q1	Attend Open Houses or parent-teacher conferences.
Sec4Q2	Attend student programs or performances.

Table 36. Parent Section 5 Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
FA	Factors	1	2	2	1	2	1
	Variance	34.5%	37.8%	40.4%	36.2%	51.5%	40.4%
RC	Alpha	0.658	0.467	0.604	0.708	0.713	0.761
	Incons.	# 1		#1	#1		
MDS	Clusters	2	2	2	2	2	2
MO H/Rho	Scale 1	0.47/0.71	0.56/0.70	0.44/0.66	0.52/0.74	0.51/0.74	0.48/0.79
	Original Scale	0.47/0.71	0.35/0.56	0.39/0.66	0.46/0.74	0.46/0.75	0.48/0.79

Most of the factor analysis models in **Table 36** explain 35%-40% of the variance with one exception – the career centers scale factor analysis which explains 51% of the variance. The overall reliability is not very good and for 3 of the scales it is even below the threshold of 0.7. The scalability is weak to moderate.

Item # 1 is inconsistent for all schools data and for elementary and middle schools.

Sec5Q1	Visit my child's classrooms during the school day.

The Rate scale again is one of the best. The one-factor factor analysis in **Table 37** explains about 70% of the variance. The reliability is very high (Alpha>0.9) and there are no inconsistent items. The scalability is very strong.

Table 37. Parent Rate Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
FA	Factors	1	1	1	1	1	1
	Variance	71.5%	71.4%	69.4%	68.0%	67.2%	65.2%
RC	Alpha	0.927	0.926	0.919	0.914	0.910	0.900
	Incons.						
MDS	Clusters	3	3	3	3	3	3
MO H/Rho	Scale 1	0.78/0.94	0.79/0.94	0.76/0.93	0.74/0.93	0.73/0.92	0.74/0.92
	Original Scale	0.78/0.94	0.79/0.94	0.76/0.93	0.74/0.93	0.73/0.92	0.74/0.92

On the opposite end is the True/False scale (**Table 38**) which is the worst of all scales. It is not reliable. A reduced small scale has strong scalability and the reliability is above the threshold of 0.7 although barely. As before we would recommend this scale to be dropped or modified into questions with Likert type answers.

Table 38. Parent True/False Scale Analysis by Type of School

Method		All Schools	Elementary Schools	Middle Schools	High Schools	Career Centers	Special Schools
RC	Alpha	0.543	0.509	0.530	0.605	0.582	0.491
	Incons.	# 4	#4	#4	#4	#4	#2,4
MDS	Clusters	3	3	3	3	3	3
MO H/Rho	Scale 1	0.61/0.74	0.55/0.72	0.63/0.73	0.64/0.74	0.58/0.70	0.60/0.71
	Original Scale	0.22/0.57	0.20/0.51	0.22/0.60	0.29/0.67	0.27/0.61	0.20/0.51

CONCLUSION

The analysis based on the combined data and by school type showed that in general the scales used in the South Carolina Card Surveys **have very good qualities**. The analysis also suggested some small changes in some of the scales and they are listed below:

Table 39. Suggested Scale Modifications

Survey	Scale	Suggested Modifications
Student	Overall School Environment	Exclude item #1
	LE	Exclude item #1
	SPE	None
	HSR	None
Teacher	Overall School Environment	Exclude items # 15, 16, 31
	LE	Exclude items 15 and 16
	SPE	Exclude item # 31
	HSR	None
Parent	Overall (All items)	Drop and replace with 2 combined and one single scale.
	Overall School Environment	Include Items LE 1-5, SPE 1-5, HSR 1-11
	Parent Activities	Include items Section 4: 1-8 and Section 5: 1-5
	LE	None
	SPE	None
	HSR	None
	School Rating by Parents	Include items Rate 1-5
	Parent Involvement (True/False)	1. Drop the scale altogether, or 2. Reduce the scale to include only items 5-7, or 3. Modify the scale by changing the questions and transforming the answers from Yes/No to a Likert-type scale.

Based on the conclusions from the scale analyses we would suggest that the total score for each scale is generated and published on the web. The score will be a simple sum of the numerical value of the valid answers. If there is at least one missing answer or answer of "Don't Know" this survey should be excluded from the score generation.

Let us for example produce a possible report on the student scale. The overall modified scale has 42 items, each with four possible valid answers: 1,2,3,4:

For students and teachers they are: 1 =Disagree 2=Mostly Disagree 3=Mostly Agree 4=Agree. For parents they are: 1 =Strongly Disagree 2=Disagree 3=Agree 4=Strongly Agree. Although slightly different they are still comparable.

This way the minimum score will be 42 and the maximum 42*4 = 168. We could also produce a proportion of the maximum score so the ideal school will have a score of 100 and the minimum will be 25 ((42/168) *100).

Table 40. School Environment Ratings for 2003/2004 School Year

Maximum score of 100

	Overall School Environment		
School	**Students**	**Teachers**	**Parents**
A	76	78	70
B	75	77	65
Etc.			

Similar tables can be produced for **Learning Environment, Social and Physical Environment, and Home-School Relations**.

In addition to the total scores for each scale a **ranking** can be produced based on the total scores. For example, based on students surveys we can calculate the total score for the overall school environment. Then, we can sort the schools by the total score and give the best school with the highest score rank 1. This will be the school in South Carolina with the most favorable school environment. The school with the next to high score will get rank 2, etc. Schools with equal scores will get the same rank.

Similar ranking could be done using the teachers and parents surveys.

Table 41. School Rankings for 2003/2004 School Year.

	Overall School Environment		
School	**Students**	**Teachers**	**Parents**
A	12	7	9
B	21	15	13
Etc.			

Similar tables can be produced for Learning Environment, Social and Physical Environment, and Home-School Relations.

From the parent survey we can calculate the total score for a few additional scales. The first is the **Parent Activity** scale which combines Sections 4 and 5. First we have to reverse the scoring because people usually expect a bigger score to mean better. For this scale the original answers are: 1 =I do this; 2=I don't do this, but I would like to; 3=I don't do this, and I don't care to. The recoding should be "old answer 1 = new answer 3", 2=2, and "old answer 3 = new answer 1." So the maximum total score for Parent Activity will be 45 (15*3) and the minimum will be 15 and the higher the score the better the parent activity in this school.

The next scale is the **School Rating** by the parents (the Rate scale). It also has to be reverse coded because the current coding is: 1 =Very good 2=Good 3=Okay 4=Bad 5=Very bad. After the recoding the bigger the total score the better the school with a minimum score of 5 and maximum of 25.

RECOMENDATIONS

1. **Modify the scales** as suggested in Table 39.
2. **Calculate the total scores for overall scale, and for LE, SPE, and HSR scales** from student, teacher, and parent surveys. An example of possible report is given in Table 42.

Table 42. School Environment Ratings by Students, Teachers, and Parents.

Maximum score of 100

	Overall School Environment		
School	**Students**	**Teachers**	**Parents**
A	76	78	70
B	75	77	65*
	Learning Environment		
School	**Students**	**Teachers**	**Parents**
A	NA**	73	65
B	70	72	60*
	Social and Physical Environment		
School	**Students**	**Teachers**	**Parents**
A	81	83	75
B	80	82	70*
	Home-School Relations		
School	**Students**	**Teachers**	**Parents**
A	74	80	NA**
B	73	80	65*

* Evaluation based on less than 30 surveys and might not be reliable.
** Insufficient number of surveys for evaluation to be valid.

The "insufficient" number in general can be about 10 surveys but it may depend on the total number of possible surveys, e.g. if a school has only 9 teachers all together.

3. **Calculate the total score for parent activity and parent involvement and the school rating by the parents**. A possible report is presented in Table 43.

Table 43. School Ratings and Evaluation by Parents.

Maximum score of 100

	School Rating
School	**Parents**
A	70
B	65*
	Parent Activities
School	**Parents**
A	65
B	60*
	Parent Involvement
School	**Parents**
A	65
B	NA**

* Evaluation based on less than 30 surveys might not be reliable.
** Insufficient number of surveys for evaluation to be valid.

4. **One item from a subscale should not be used as a substitute for the total score** of a scale. For example, item # 18 from student survey Learning Environment scale should not be used instead of the total score based on all 17 items (2-18).

Our analysis showed that the full scales and subscales have very good scalability and reliability which means that they should be used as scales, i.e. by calculating total scale scores. In addition in some cases the most difficult items were inconsistent with the other items of the scale. It is not fair for example, the whole learning environment to be judged just by one question disregarding the other 16 questions.

If the current practice of using only one item per scale is very desirable, the practice could continue but the total scores for all the scales should also be available in addition to the single question.

5. **Sample size of 30 should be used as a guideline** for including the results for each school. In other words if there are less than 30 responses from a school the total scores and ranks should be either not published or published with a footnote saying that less than 30 surveys were received from this school and the result may not be reliable. Schools with less than 10 valid surveys should not get total score.

6. **The surveys should be filled out correctly and completely** which would reduce the missing data problem significantly. Schools that receive less than **30 valid** surveys for any of the 3 groups: student, teacher and parent surveys should be contacted and urged to work more to increase the number of surveys received and work particularly with parents, and Elementary and Middle school students.

7. **Additional statistical techniques can be used to fill in the missing data** if the sufficient number of valid surveys per school remains a problem**.** For example, when just a few items are missing from a scale they can be replaced by the means for this scale. In complicated cases more rigorous statistical approach like multiple imputations could be applied. The bottom line is to avoid at all cost giving an evaluation of a school based only on a few surveys as opposed to other school that has 300 responses.

FURTHER ANALYSIS

After establishing the quality of the scales and being able to compute the total scores for each scale and subscales the next step is to analyze the results from the SC Report Card surveys. We think that the following analyses will enrich the understanding of the school environment and its factors.

1. Producing an annual report.

Annually produce the total scores for the 3 overall scales for students, teachers and parents, and the 3 main subscales: Learning Environment, Social and Physical Environment, and Home-School Relations. Also, from the parent surveys, present the 2 additional scales: Parents Activities, and Parents School Ratings. This would require aggregating the data files by BEDS (School ID), creating the scale scores and counting the number of surveys from student, teachers, and parents for each school. The information based on less than 30 surveys per school should be footnoted as possibly not reliable. Certainly, if the number of surveys for a particular school is very low (e.g. less than 10) the scores should not be produced. These schools should be informed of the lack of surveys and they should encourage their students, teachers and parents to participate.

2. Analyzing the quality of the data.

Our **preliminary studies** on a limited basis show that from the valid student surveys, i.e. with no missing data, about 10-12% of the schools have less than 30 valid surveys. On the other hand some schools have as high as 472 valid student surveys. The lack of enough surveys seems to be disproportionately distributed by school type. High schools have the greatest number of valid student surveys. Elementary, Middle school and Special schools have very few valid surveys.

There are problems with the teacher surveys as well. More than 30% of the schools have less than 30 valid teachers' surveys per school. On the other hand there does not seem to be a difference by school type as was the case with the students.

For the parent surveys the situation is also difficult. More than 50% of the schools have less than 30 parents' surveys available per school. In contrast to the student surveys here the majority of the surveys come from parents of Elementary school students not High School students. These differences need to be addressed and further investigated.

If the sufficient number of valid surveys per schools remains a problem additional techniques might be used to fill in the missing data**.** For example, when just a few items are missing from a scale they can be replaced by the means for this scale. In more complicated cases more rigorous statistical approach like multiple imputations could be applied.

3. Analysis of the data from SC Report Card Surveys.

Having available the total scores for different scales the next logical thing is to analyze the results. Some of the possible research questions here are as follows:

1. What are the average scores by type of school for the different scales?
2. Are there any relationships between student, teacher, and parent survey results?
3. Are there any differences in the school environment by school type?
4. Are there any relationships between the overall scales and between the subscales?
5. Are there any differences between school districts?

Aggregating the results by school districts should be done carefully and additional factors might be considered relevant and important in explaining possible differences. Such factors included Urban/Rural setting, school size, etc. and other clusters deemed relevant.

Some *preliminary* results based only on the valid surveys are presented below.

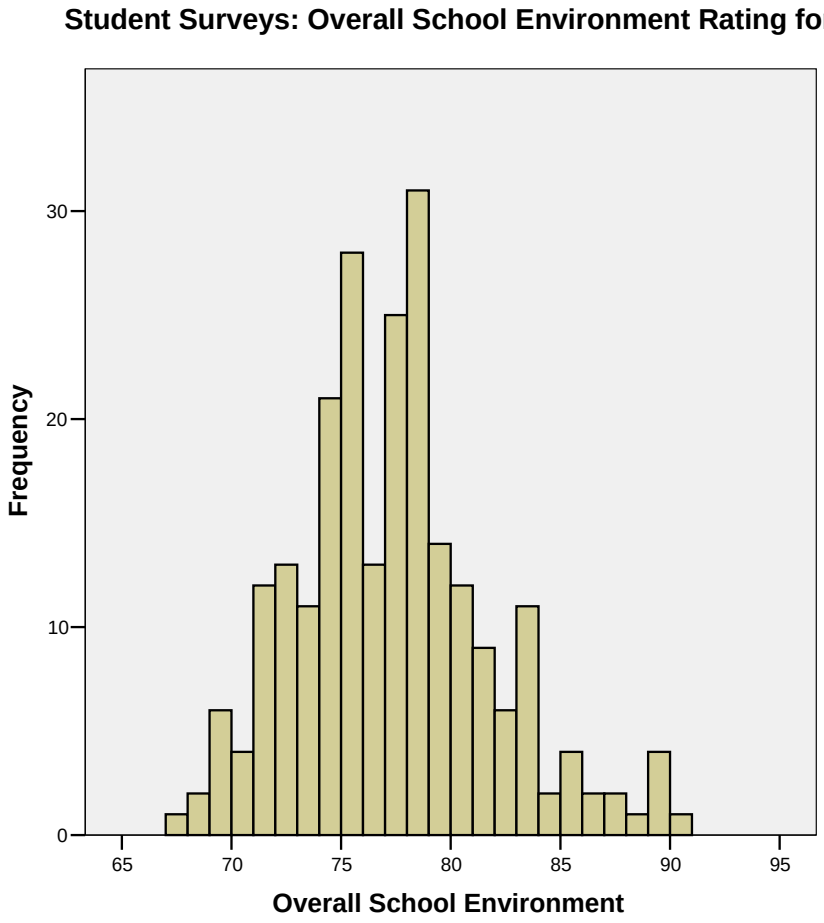
Table 44. SC Report Card Surveys Results for 2003/4 School Year

Maximum score of 100

	Overall School Environment		
School	**Students**	**Teachers**	**Parents**
A	80.3	82.7	73.0
B	74.6	82.3	68.8
C	79.7	83.9	71.8
D	73.3	74.7	63.8
Etc.			

These are actual ratings for some of the best schools with 3 valid overall scales.

Figure 1. Overall School Environment Rating Distribution



4. Analysis of school environment and school performance.

For this analysis the aggregated scale files have to be merged with the school performance files. The later contain very valuable information about the students, teachers and the school: overall school rating, number of students, students' GPA, graduation rates, student/teacher ratio, teachers with degree, returning teachers, etc. These files also contain some of the scale questions but as we showed in the previous analysis this information is inadequate and it should be replaced by the total scores and only from valid surveys.

This is very valuable information by itself but additional questions might be interesting to answer.

1. What are the factors that determine the school ranking? What is their influence?

2. What are the differences between schools with high ranking and schools with low ranking regarding the school environment?
3. What are the differences between schools with high ranking and schools with low ranking regarding the students, teachers and other school characteristics?
4. Are there any relationships between the students' characteristics and the students' evaluation of the school environment?
5. Are there any relationships between the teachers' characteristics and the teachers' evaluation of the school environment?

5. Analysis of the dynamics of the school environment and school performance.

In order to perform this analysis a new database has to be created with the SC Card Report surveys and school performance data for the last several years if possible. At least 5 years' data are necessary for meaningful analysis of the dynamics.

The relevant research questions here are as follows:

1. Are there any significant trends for the overall school environment and school performance? Are the performance and environment improving over time?
2. Are there any differences in the trends for school environment and school performance?
3. Is there any lag in the performance results in relation to school environment? For example, is it the case that first the environment improves and it later leads to better performance? How many years is the lag, if any?
4. Is there any lag in the performance results in relation to teachers' characteristics? For example, is it the case that first the teachers improve (more teachers with degree, more experience, etc.) which later leads to better performance?
5. Are there any differences in dynamics for different school districts?

REFERENCES

Cox, T.F. and Cox, M. A. A. (2001) Multidimensional Scaling. Chapman & Hall.

DeVillis, R. (2003). Scale Development. Theory and Applications.

Habing, B. (2005). Course Notes, http://www.stat.sc.edu/~habing/courses/pastcourses.html, USC, Department of Statistics.

Ripley, B. D. (1996) Pattern Recognition and Neural Networks, Cambridge University Press.

Sijtsma, K., Molenaar, I. (2002). Introduction to Nonparametric Item response Theory.

Stout, W., Habing, B., Douglas, J. , Kim, H.R. , Roussos, L. & Zhang, J. (1996). Conditional Covariance-Based Nonparametric Miltidimensionality Assessment, Applied Psychological Measurement, Vol. 20, No. 4, December 1996, pp. 331-354.

van Abswoude, A., van der Ark, L., Sijtsma, K. (2004). A Comparative Study of Test Data Dimensionality Assessment Procedures Under Nonparametric IRT Models, Applied Psychological Measurement, Vol. 28, No.1, pp. 3-24.

Venables, W. N. and Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth edition. Springer.

SOFTWARE

1. HCA/CCPROX, Assessment Systems Corporation (www.assess.com).
2. MSPWIN ver. 5.0, ProGAMMA, www.gamma.rug.nl
3. SPSS ver. 13
4. SAS ver. 9
5. R ver. 2.1.0