Abstract
Reliability refers to accuracy and precision of a measurement instrument. Confirmatory factor analysis (CFA) is a statistical technique used to verify the factor structure of a measurement instrument. EFA, traditionally, is used to explore the possible underlying factor structure of a measurement instrument.

Guidelines for reliability, confirmatory and exploratory factor analysis will be discussed. Examples using the Race and Schooling Instrument (Revised, Shay 2008) and PROC CORR, PROC CALIS and PROC FACTOR illustrate reliability, CFA and EFA statistical techniques.

Introduction
CFA and EFA are powerful statistical techniques. An example of CFA and EFA could occur with the development of measurement instruments, e.g. a satisfaction scale, attitudes toward health, customer service questionnaire. During the process, a blueprint is developed, questions written, a response scale determined, the instrument pilot tested, data collected, and CFA completed. The blueprint identifies the factor structure or what we think it is. If questions do not measure what we thought they should, the factor structure does not follow our blueprint, the factor structure is not confirmed, and EFA is the next step. EFA helps us determine what the factor structure looks like according to participant responses. Exploratory factor analysis is essential to determine underlying constructs for a set of measured variables.

Measurement Instrument
The example for this presentation analyzes data collected with the Race and Schooling School-Based Instrument (Shay, 2008). The 23-item instrument measures the essential principles of culturally responsive teaching and learning. Each question, aligning to a principle and category, provides insight into participants’ levels of involvement in culturally responsive teaching practices and attitudes and beliefs toward confronting institutional bias and discrimination in schools (Shay, in press).

The Race and Schooling School-Based Instrument was developed by Singleton (2003) to include seven categories. Permission to use and adjust the instrument for research purposes was granted by Singleton to Shay (2008). See Appendix A to review the survey instrument.

The original survey instrument had a 5-point Likert scale with (1) Rarely, (3) Sometimes, and (5) Often as response choices to survey questions. In the revised version, responses were changed to (1) Almost Never; (2) Seldom; (3) Sometimes; (4) Frequently; and (5) Almost Always. Additional questions were added to allow respondents to provide information on demographics, e.g., years of experience in the district and in the field of education.

Data was collected in a metropolitan School District in Midwestern United States. Of the 282 respondents, 16% were male (n=42) and 84% were female (n=220). The majority of respondents were white (89.1%, n=229) with 3.5% Hispanic (n=9), 3.5% Black (n=9), 2.0% Asian (n=5) and 2.0% identified their race as Other.

Reliability
Reliability refers to the accuracy and precision of a measurement procedure (Thorndike, Cunningham, Thorndike, & Hagen, 1991). Reliability may be viewed as an instrument’s relative lack of error. In addition, reliability is a function of properties of the underlying construct being measured, the test itself, the groups being assessed, the testing environment, and the purpose of assessment. Reliability answers the question, How well does the instrument measure what it purports to measure?

Some degree of inconsistency is present in all measurement procedures. The variability in a set of item scores is due to the actual variation across individuals in the phenomenon that the scale measures, made up of true score and error. Therefore, each observation of a measurement (X) is equal to true score (T) plus measurement error (e), or X = T + e. Another way to think about total variation is that it has two components: “signal” (i.e., true differences in the latent construct) and “noise” (i.e., differences caused by everything but true differences).
Sources of measurement inconsistency could be due to

1) a person changing from one testing to the next
   a. the amount of time between tests may have resulted in growth or change
   b. motivation to perform may be different at each testing
   c. the individual may be physically better able to perform, e.g., more rested
   d. the individual may have received tutoring between testings

2) the task being different from one testing to the next
   a. different environment
   b. different administrator
   c. different test items on parallel forms

3) the sample of behavior resulting in an unstable or undependable score
   a. the sample of behavior and evaluation of it are subject to chance or random influences
   b. a small sample of behavior does not provide a stable and dependable characterization of an individual
   c. for example, the average distance of 100 throws of a football would provide a more stable and accurate
      index of ability than a single throw.

Reliability may be expressed

1) as an individual’s position in a group (correlation between first and second measurements; the more nearly
   individuals are ranked in the same order, the higher the correlation and the more reliable the test)

2) within a set of repeated measures for an individual (internal consistency, how consistency are items answered)

Reliability can be assessed by

1) repeating the same test or measure (test-retest)
2) administering an equivalent form (parallel test forms)
3) using single-administration methods
   a. subdividing the test into two or more equivalent parts
   b. internal consistency – measured with Cronbach’s coefficient alpha.

Internal Consistency

Internal consistency is a procedure to estimate the reliability of a test from a single administration of a single form.
Internal consistency depends on the individual’s performance from item to item based on the standard deviation of
the test and the standard deviations of the items.

\[
\gamma = \frac{\left( \frac{1}{n} \sum_{i=1}^{n} \left( SD_i^2 - \bar{SD}^2 \right) \right)}{\left( \frac{1}{n-1} \sum_{i=1}^{n} SD_i^2 \right)}
\]  

(1)

where \( \gamma \) is the estimate of reliability,
   n is the number of items in the test,
   \( SD_i \) is the standard deviation of the test scores
   \( \bar{SD} \) means “take the sum” and covers the n items,
   \( SD_i \) is the standard deviation of the scores from a group of individuals on an item.

KR20, Kuder-Richardson Formula 20, is a special form of coefficient alpha that applies when items are dichotomous
(e.g., yes/no, true/false) or are scored as right or wrong.

Factors Influencing Reliability

Many factors can affect the reliability of a measurement instrument. They are the

1) range of the group
   a. pooling a wider range of grades or ages produces a reliability coefficient of higher magnitude
   b. take into account the sample on which the reliability coefficient is based when comparing
      instruments

2) level of ability in the group
   a. precision of measurement can be related to ability level of the people being measured
   b. report the standard error of measurement for different score levels

3) methods used for estimating reliability
   a. amount of time between administrations
   b. method of calculating reliability
4) length of the test
   a. when reliability is moderately high, it takes a considerable increase in test length to increase reliability
   b. relationship of reliability to length of test can be expressed with

\[ r_{kk} = \frac{k r_0}{1 + (k - 1) r_0} \]  

where \( r_{kk} \) is the reliability of the test \( k \) times as long as the original test,
\( r_0 \) is the reliability of the original test, and
\( k \) is the factor by which the length of the test is changed.

For example, if reliability is .60 for a 10-item instrument, what is reliability for a 20-item instrument?
\[ r_{kk} = \frac{2(.60)}{1 + (2 - 1)(.60)} = \frac{1.20}{1.60} = 0.75 \]

Levels of Reliability
Acceptable levels of reliability depend on the purpose of the instrument. Acceptable reliability of instruments developed for research purposes can be as low as 0.60. An acceptable reliability level of a diagnostic instrument used for making decisions about individuals (e.g., a psychological measure) should be much higher, e.g., 0.95.

Comparisons
The reliability coefficient provides a basis for assessment instrument comparison when measurement is expressed in different scales. The assessment with the higher reliability coefficient could provide a more consistent measurement of individuals.

Statistical Power
An often overlooked benefit of more reliable scales is that they increase statistical power for a given sample size (or allow smaller sample size to yield equivalent power), relative to less reliable measures. A reliable measure, like a larger sample, contributes relatively less error to the statistical analysis. The power gained from improving reliability depends on a number of factors including
1) the initial sample size
2) the probability level set for detecting a Type I error
3) the effect size (e.g., mean difference) that is considered significant
4) the proportion of error variance that is attributable to measure unreliability rather than sample heterogeneity or other sources.

To raise the power, substitute a highly reliable scale for a substantially poorer one. For example, when \( n = 50 \), two scales with reliabilities of 0.38 have a correlation of 0.24, \( p < 0.10 \), and would be significant at \( p < 0.01 \) if their reliabilities were increased to 0.90 or if the sample was more than twice as large (\( n > 100 \)).

PROC CORR and options for Reliability
DATA= input data set
OUTP= output data set with Pearson correlation statistics
ALPHA compute Cronbach’s coefficient alpha
NOMISS exclude observations with missing analysis values
NOCORR suppresses printing Pearson correlations
NOSIMPLE suppresses printing descriptive statistics

Confirmatory Factor Analysis
CFA allows the researcher to test the hypothesis that a relationship between the observed variables and their underlying latent construct(s) exists. The researcher uses knowledge of the theory, empirical research, or both, postulates the relationship pattern a priori and then tests the hypothesis statistically.
Exploratory Factor Analysis

Psychologists searching for a neat and tidy description of human intellectual abilities lead to the development of factor analytic methods. Galton, a scientist during the 19th and 20th centuries, laid the foundations for factor analytic methods by developing quantitative methods to determine the interdependence between 2 variables. Karl Pearson was the first to explicitly define factor analysis. In 1902, Macdonnell was the first to publish an application of factor analysis. His study compared physical characteristics of 3000 criminals and 1000 Cambridge undergraduates.

Factor analysis could be described as orderly simplification of interrelated measures. Traditionally factor analysis has been used to explore the possible underlying structure of a set of interrelated variables without imposing any preconceived structure on the outcome (Child, 1990). By performing exploratory factor analysis (EFA), the number of constructs and the underlying factor structure are identified.

Goals of factor analysis are
1) to help an investigator determine the number of latent constructs underlying a set of items (variables)
2) to provide a means of explaining variation among variables (items) using few newly created variables (factors), e.g., condensing information
3) to define the content or meaning of factors, e.g., latent constructs

Assumptions underlying EFA are
• Interval or ratio level of measurement
• Random sampling
• Relationship between observed variables is linear
• A normal distribution (each observed variable)
• A bivariate normal distribution (each pair of observed variables)
• Multivariate normality

Limitations of EFA are
• the correlations, the basis of factor analysis, describe relationships. No causal inferences can be made from correlations alone.
• the reliability of measurement instrument (avoid instrument with low reliability)
• sample size (larger sample \(\rightarrow\) larger correlation)
  o minimal number for reliable results is greater than 100 and 5 times the number of items
  o since some subjects may not answer every item, a larger sample is desirable, e.g., for 30 items, at least 150 subjects \((5 \times 30)\), a sample of 200 subjects would allow for missing data.
• sample selection
  o Representative of population
  o Do not pool populations
• variables could be sample specific, e.g., a unique quality possessed by a group does not generalize to the population
• nonnormal distribution of data
**Factor Extraction**

Factor analysis seeks to discover common factors. The technique for extracting factors attempts to take out as much common variance as possible in the first factor. Subsequent factors are, in turn, intended to account for the maximum amount of the remaining common variance until, hopefully, no common variance remains.

Direct extraction methods obtain the factor matrix directly from the correlation matrix by application of specified mathematical models. Most factor analysts agree that direct solutions are not sufficient. Adjustment to the frames of reference by rotation methods improves the interpretation of factor loadings by reducing some of the ambiguities which accompany the preliminary analysis (Child, 1990). The process of manipulating the reference axes is known as rotation. The results of rotation methods are sometimes referred to as derived solution because they are obtained as a second stage from the results of direct solutions.

Rotation applied to the reference axes means the axes are turned about the origin until some alternative position has been reached. The simplest case is when the axes are held at 90° to each other, orthogonal rotation. Rotating the axes through different angles gives an oblique rotation (not at 90° to each other).

**Methods**

As an aside, names given to factor extraction methods have some interesting origins.

- **Procrustes** was a highwayman who tied his victims to a bed and shaped them to its structure either by stretching them or by cutting off their limbs. In factor analysis, the Procrustes technique/method involves testing data to see how close they fit a hypothesized factor structure.
- **The plasmode method** is taken from well-established areas (e.g., physics, chemistry) so that the factor structure is predictable.

**Criteria for Extracting Factors**

Determining the number of factors to extract in a factor analytic procedure is dependent on meeting appropriate criteria. They are

1) **Kaiser’s criterion**, suggested by Guttman and adapted by Kaiser, considers factors with an eigenvalue greater than one as common factors (Nunnally, 1978)
2) **Cattell’s (1966) scree test**. The name is based on an analogy between the debris, called scree, that collects at the bottom of a hill after a landslide, and the relatively meaningless factors that result from overextraction. On a scree plot, because each factor explains less variance than the preceding factors, an imaginary line connecting the markers for successive factors generally runs from top left of the graph to the bottom right. If there is a point below which factors explain relatively little variance and above which they explain substantially more, this usually appears as an “elbow” in the plot. This plot bears some physical resemblance to the profile of a hillside. The portion beyond the elbow corresponds to the rubble, or scree, that gathers. Cattell’s guidelines call for retaining factors above the elbow and rejecting those below it. This amounts to keeping the factors that contribute most to the variance
3) **Proportion of variance accounted for** keeps a factor if it accounts for a predetermined amount of the variance (e.g., 5%, 10%).
4) **Interpretability criteria**
   a. Are there at least 3 items with significant loadings (>0.30)?
   b. Do the variables that load on a factor share some conceptual meaning?
   c. Do the variables that load on different factors seem to measure different constructs?
   d. Does the rotated factor pattern demonstrate simple structure? Are there relatively
      i. high loadings on one factor?
      ii. low loadings on other factors?

**Statistics**

EFA decomposes an adjusted correlation matrix. Variables are standardized in EFA, e.g., mean=0, standard deviation=1, diagonals are adjusted for unique factors, 1-u. The amount of variance explained is equal to the trace of the matrix, the sum of the adjusted diagonals or communalities. Squared multiple correlations (SMC) are used as communality estimates on the diagonals. Observed variables are a linear combination of the underlying and unique factors. Factors are estimated, \((X_1 = b_{1F1} + b_{2F2} + \ldots + e_1)\) where \(e_1\) is a unique factor).

Factors account for common variance in a data set. The amount of variance explained is the trace (sum of the diagonals) of the decomposed adjusted correlation matrix. Eigenvalues indicate the amount of variance explained by each factor. Eigenvectors are the weights that could be used to calculate factor scores. In common practice, factor scores are calculated with a mean or sum of measured variables that “load” on a factor.
The EFA Model is \( Y = X\beta + E \)

where
- \( Y \) is a matrix of measured variables
- \( X \) is a matrix of common factors
- \( \beta \) is a matrix of weights (factor loadings)
- \( E \) is a matrix of unique factors, error variation

Communality is the variance of observed variables accounted for by a common factor. A large communality value indicates a strong influence by an underlying construct. Community is computed by summing squares of factor loadings

\[ d_1^2 = 1 - \text{communality} = \% \text{ variance accounted for by the unique factor} \]

\[ d_1 = \sqrt{1-\text{community}} = \text{unique factor weight (parameter estimate)} \]

**EFA Steps**

1) initial extraction
   - each factor accounts for a maximum amount of variance that has not previously been accounted for by the other factors
   - factors are uncorrelated
   - eigenvalues represent amount of variance accounted for by each factor

2) determine number of factors to retain
   - scree test, look for elbow
   - proportion of variance
   - prior communality estimates are not perfectly accurate, cumulative proportion must equal 100% so some eigenvalues will be negative after factors are extracted, e.g., if 2 factors are extracted, cumulative proportion equals 100% with 6 items, then 4 items have negative eigenvalues
   - interpretability
     - at least 3 observed variables per factor with significant factors
     - common conceptual meaning
     - measure different constructs
     - rotated factor pattern has simple structure (no cross loadings)

3) rotation – a transformation

4) interpret solution

5) calculate factor scores

6) results in a table

7) prepare results, paper

**PROC FACTOR and options for EFA**

- **DATA =** specifies dataset to be analyzed
- **PRIORS =** specifies prior communality estimate
- **METHOD =** specifies maximum likelihood and unweighted least squares methods
- **ROTATE =** specifies rotation method
- **SCREE** requests a scree plot of the eigenvalues
- **N =** specifies number of factors
- **MINEIGEN=** specifies select factors with eigenvalues greater than 1
- **OUT =** data and estimated factor scores, use raw data and N=
- **FLAG =** include a flag (*) for factor loadings above a specified value
- **REORDER =** arranges factor loadings from largest to smallest for each factor

An example of SAS code to run EFA. priors specify the prior communality estimate

```sas
proc factor method=ml priors=smc
   maximum likelihood factor analysis
method=uls priors=smc
   unweighted least squares factor analysis
method=prin priors=smc
   principal factor analysis.
```
**Similarities between CFA and EFA**
- Both techniques are based on linear statistical models.
- Statistical tests associated with both methods are valid if certain assumptions are met.
- Both techniques assume a normal distribution.
- Both incorporate measured variables and latent constructs.

**Differences between CFA and EFA**

**CFA** requires specification of
- a model a priori
- the number of factors
- which items load on each factor
- a model supported by theory or previous research
- error explicitly

**EFA**
- determines the factor structure (model)
- explains a maximum amount of variance

**Statistical Analysis**
With background knowledge of reliability, exploratory and confirmatory factor analysis, we’re ready to proceed to the statistical analysis!

**Reliability Analysis with PROC CORR (7 factors)**
Data was analyzed with PROC CORR to determine the degree of internal consistency (reliability). A Cronbach alpha statistic indicates the level of reliability. Values could range from 0.0 to 1.0 with a value closer to 1.0 indicating a higher level of reliability.

Reliability ranged from 0.62 to 0.88 on the total and category subscales for the sample data on the Race and Schooling School-Based Instrument. For research purposes, reliability was acceptable (see Table 1).

**Table 1. The Race and Schooling School-Based Instrument - Reliability**

<table>
<thead>
<tr>
<th>Factor (category)</th>
<th>Items</th>
<th>Alpha (Standardized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional Development</td>
<td>1, 2, 3, 4, 5</td>
<td>0.80</td>
</tr>
<tr>
<td>Relationships</td>
<td>6, 7, 8</td>
<td>0.63</td>
</tr>
<tr>
<td>Engagement</td>
<td>9, 10, 11</td>
<td>0.70</td>
</tr>
<tr>
<td>Learning and Teaching</td>
<td>12, 13, 14</td>
<td>0.73</td>
</tr>
<tr>
<td>Achievement Results</td>
<td>15, 16, 17</td>
<td>0.62</td>
</tr>
<tr>
<td>Community</td>
<td>18, 19, 20</td>
<td>0.63</td>
</tr>
<tr>
<td>School Culture</td>
<td>21, 22, 23</td>
<td>0.79</td>
</tr>
<tr>
<td>Total</td>
<td>1-23</td>
<td>0.88</td>
</tr>
</tbody>
</table>
SAS Code

PROC CORR procedure calculates a Cronbach Alpha statistics for the reliability analysis. The following code calculates Cronbach Alpha for the total scale (questions 1-23) and for each category subscale.

```sas
proc corr data=rawsub2 nocorr alpha nomiss;
  var q1-q23;
proc corr data=rawsub2 nocorr alpha nomiss;
  var q1-q5;
proc corr data=rawsub2 nocorr alpha nomiss;
  var q6-q8;
proc corr data=rawsub2 nocorr alpha nomiss;
  var q9-q11;
proc corr data=rawsub2 nocorr alpha nomiss;
  var q12-q14;
proc corr data=rawsub2 nocorr alpha nomiss;
  var q15-q17;
proc corr data=rawsub2 nocorr alpha nomiss;
  var q18-q20;
proc corr data=rawsub2 nocorr alpha nomiss;
  var q21-q23;
```

Confirmatory Factor Analysis (7 factor)

Category subscale scores (factors) were created with the MEAN function. The factors were subjected to a confirmatory factor analysis (CFA) with PROC CALIS. Figure 1 illustrates the CFA model. The SAS code below tests the underlying factor structure.

```sas
proc calis data=rawsub2 kurtosis;
  lineqs
    Professional_Development=s1 F1 + e1,
    Relationships        =s2 F1 + e2,
    Engagement           =s3 F1 + e3,
    Learning_Teaching    =s4 F1 + e4,
    AchResults           =s5 F1 + e5,
    Community            =s6 F1 + e6,
    SchoolCulture        =s7 F1 + e7;
  std
    e1-e7=vare1-vare7,
    F1=1;
  var Professional_Development Relationships Engagement Learning_Teaching AchResults Community SchoolCulture;
```

Another step that could be taken in the confirmatory approach is to test the structure for each of the 7 factors. The following SAS code would test a model for the Professional Development factor.

```sas
proc calis data=rawsub2 kurtosis; **modification;
  lineqs
    q1= p1 F1 + e1,
    q2= p2 F1 + e2,
    q3= p3 F1 + e3,
    q4= p4 F1 + e4,
    q5= p5 F1 + e5;
  std
    e1-e5=vare1-vare5,
    F1=1;
  var q1-q5;
```

If each factor can be confirmed, then a combined model is evaluated. If this model is confirmed, then a model including one latent construct (race and schooling) with a direct relationship to each factor is evaluated. Each factor influences responses to items. For example, the latent construct professional development influences responses to items 1, 2, 3, 4 and 5.
PROC CALIS procedure provides the number of observations, variables, estimated parameters, and informations (related to model specification), descriptive statistics, and multivariate kurtosis. PROC CALIS procedure also indicates the observations with the largest contribution to kurtosis (see Figure 2).

**Fit Statistics**
Determine criteria a priori to access model fit and confirm the factor structure. Some of the criteria indicate acceptable model fit while other are close to meeting values for acceptable fit (see Figure 3).

- Chi-square describes similarity of the observed and expected matrices. Acceptable model fit is indicated by a chi-square probability greater than or equal to 0.05. For this CFA model, the chi-square value is not close to zero (Chi-square=35.7367) and $p = 0.0011$, not greater than 0.05.
- RMSEA indicates the amount of unexplained variance or residual. A value of 0.743 RMSEA value is larger than the 0.06 or less criteria.
- CFI (0.9472), NNI (0.9208), and NFI (0.9174) values meet the criteria (0.90 or larger) for acceptable model fit.
For purposes of this example, 3 fit statistics indicate acceptable fit and 2 fit statistics indicate unacceptable fit. The CFA analysis has not confirmed the factor structure. If the analysis indicates unacceptable model fit, the factor structure cannot be confirmed, an exploratory factor analysis is the next step.

The factor structure is not confirmed. No further investigation of the confirmatory model is necessary, parameter estimates, variances, covariances. Proceed with exploratory factor analysis to determine the factor structure.
Exploratory factor analysis with PROC FACTOR

- method is maximum likelihood
- diagonals of the correlation matrix are equal to squared multiple correlations
- criteria set a priori is each factor retained will explain at least 10% of the variance

SAS Code

```sas
proc factor data=rawsub2 method=ml priors=smc rotate=varimax reorder;
var q1-q23;
```

Results show 5 factors. Three factors each explain at least 10% of the variance (59.67%, 19.11%, 11.19%) while 2 factors each explain less than 10% of the variance (8.86%, 5.37%). According to criteria set a priori, each factor retained will explain at least 10% of the variance, three factors will be retained. The analysis will be rerun specifying three factors.

```sas
proc factor data=rawsub2 method=ml priors=smc rotate=varimax reorder n=3;
var q1-q23;
```

Figure 5 indicates that both hypothesis tests are rejected, no common factors and 3 factors are sufficient. In practice, we want to reject the first hypotheses and accept the second hypothesis.

Tucker and Lewis’s Reliability Coefficient indicates good reliability. Reliability is a value between 0 and 1 with a larger value indicating better reliability.
Results show items 21, 22, 23 as a factor, items 1, 2, 3, 4 as a factor, and items 5-20 as a factor. Two of the factors are the same or close to the original factor structure. Items 21, 22, 23 are the School Culture factor. Items 1, 2, 3, 4, 5 are the Professional Development factor in the original factor structure. Item 5 is not included in the revised factor structure.

**Interpretability**
Is there some conceptual meaning for each factor? Could the factors be given a name?
Factor1 could be called Professional Development (items 1, 2, 3, 4).
Factor2 could be called Attitudes (items 5-20).
Factor3 could be called School Culture (items 21, 22, 23).
Reliability with PROC CORR tested the reliability for the revised factor structure. The SAS code for reliability is

```sas
proc corr data=rawsub2 nocorr alpha nomiss;
  var q1-q4;
proc corr data=rawsub2 nocorr alpha nomiss;
  var q5-q20;
proc corr data=rawsub2 nocorr alpha nomiss;
  var q21-q23;
```

Reliability with Cronbach Alpha for items 21, 22, 23 is 0.79, for items 1, 2, 3, 4 is 0.79, and for items 5-20 is 0.88. The range is 0.79 to 0.88 for the revised factor structure. The range for the original factor structure is 0.62 to 0.80.

**Conclusion**

Confirmatory and Exploratory Factor Analysis are powerful statistical techniques. The techniques have similarities and differences. Determine the type of analysis a priori to answer research questions and maximize your knowledge.

**WAM** (Walk away message)

Select CFA to verify the factor structure and EFA to determine the factor structure.

**References**


### Professional Staff Development

<table>
<thead>
<tr>
<th>Question</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To what extent do teachers and administrators in your school talk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>openly and constructively about race with each other?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2. To what extent do staff development activities help educators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>understand the ways in which race influences student behavior?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3. To what extent do staff development activities help educators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acquire knowledge about the history and culture of various racial</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>groups?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. To what extent do staff development activities help educators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>become knowledgeable about the diverse racial perspectives on</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>historical and current events?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. To what extent are successful efforts being made in your school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to attract, retain, and advance teachers and administrators of color?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

### Relationships

<table>
<thead>
<tr>
<th>Question</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. To what extent do teachers in your school talk openly and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>constructively about race with students?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>7. To what extent do teachers encourage students to have open and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>constructive conversations about being victims of racial discrimination</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>and about possessing racial power and white privilege?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. To what extent do teachers structure interracial cooperative groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>that enable students of different racial groups to become acquainted</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>with each other?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Engagement

<table>
<thead>
<tr>
<th>Question</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. To what extent is decision-making in the school widely shared</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>among administrators, teachers, parents and students of a variety of</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>racial groups?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. To what extent do teachers encourage students of different racial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>groups to talk with each other openly and constructively about race?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>11. To what extent are deliberate actions taken by administrators and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>teachers to insure that students of color are represented proportionally</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>in extra-curricular activities and leadership roles?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Learning and Teaching

<table>
<thead>
<tr>
<th>Question</th>
<th>1 - Almost Never</th>
<th>2 - Seldom</th>
<th>3 - Sometimes</th>
<th>4 - Frequently</th>
<th>5 - Almost Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>12. To what extent do teachers help students acquire the knowledge and skills needed to have thoughtful, constructive, heartfelt discussions about race?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>13. To what extent are students provided factual information in social studies and other subject areas that contradicts misconceptions about people of color?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>14. To what extent does the school curriculum include a focus on racial power and white privilege through examples in history, art, science, and other disciplines?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

### Achievement Results

<table>
<thead>
<tr>
<th>Question</th>
<th>1 - Almost Never</th>
<th>2 - Seldom</th>
<th>3 - Sometimes</th>
<th>4 - Frequently</th>
<th>5 - Almost Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>15. To what extent do teachers use a variety of assessment devices to ensure that students of all racial groups meet rigorous standards across the curriculum?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>16. To what extent do teachers use a variety of assessment devices to measure improved race relations between students of different racial groups?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>17. To what extent are students of all racial groups consistently exposed to and supported in the most rigorous curricular opportunities?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

### Community

<table>
<thead>
<tr>
<th>Question</th>
<th>1 - Almost Never</th>
<th>2 - Seldom</th>
<th>3 - Sometimes</th>
<th>4 - Frequently</th>
<th>5 - Almost Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>18. To what extent do teachers have regular personal contact (e.g. phone, face to face) with Black and Hispanic families to advise them on the achievement of their children?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>19. To what extent are Black and Hispanic children specifically invited to become a part of school-wide activities, committees, and councils?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>20. To what extent are families of Black and Hispanic families visible in school and leadership positions?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

### School Culture

<table>
<thead>
<tr>
<th>Question</th>
<th>1 - Almost Never</th>
<th>2 - Seldom</th>
<th>3 - Sometimes</th>
<th>4 - Frequently</th>
<th>5 - Almost Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>21. To what extent do you believe that Black and Hispanic students should perform at least as well as White students?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>22. To what extent do you believe it is our responsibility as educators to make this level of achievement occur for our Black and Hispanic students?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>23. To what extent do you believe that our Black and Hispanic students can actually reach this level of achievement?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Note: Information questions are not shown in Appendix A.