

ANALYTIC STRATEGIES EMPLOYED FOR PRIDE SURVEYS LEARNING ENVIRONMENT SURVEYS

Analysis of the LES Pilot Instrument

A. Preparation of the Data Set

A total of 201,247 scanable survey forms were returned from Alabama students.

The purpose of the pilot survey was to analyze item performance for reliability. Factor analysis was the primary analytic strategy used to reach this goal. Prior to commencing the analysis, two steps were taken to prepare the data set for use.

1. Some survey items were reverse coded so that later scaling would be consistent in direction and interpretation across all of the items.
2. Various demographic variables were recoded as needed in preparation for later analyses.

B. Missing Data Analyses

The missingness status of all survey items was determined, which is displayed in Table 1. Overall, low levels of missingness were found. However, the data were positively skewed, with a relatively small number of individual items accounting for a disproportionate amount of missingness.

Table 1. Missing data in the survey

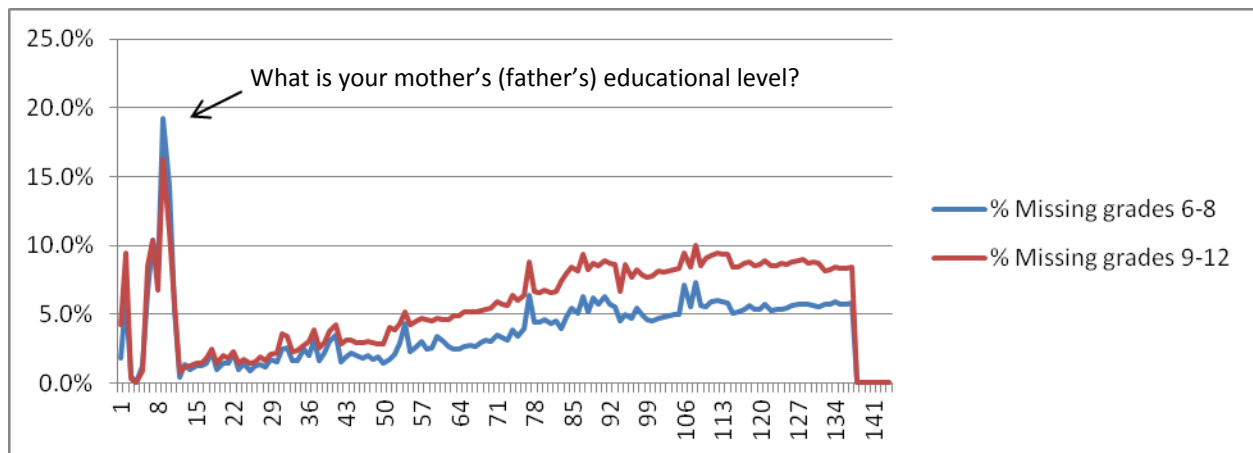
Mean number of missing items per student	6.2%
Mean percentage of items missing per student	3.2%
Median number of missing items per student	1.00

The specific missingness problem is best illustrated in Figure 1. In this figure, the survey items are arrayed left to right in order of presentation. The vertical axis codes for the percentage of survey forms that had a missing value on each survey item.

Across all figures a slight, very gradual increase in the overall level of missingness can be observed. In other words, students became slightly more likely to not answer a survey question as the survey progressed. However, this effect is quite small and does not impact the overall integrity of the LES. Thus, we found no evidence of significant respondent fatigue with the LES in the pilot form format.

High levels of missingness were associated with the two items related to parents' educational level. Figure 1 also shows that, in general, being in grades 9-12 is associated with higher missingness than being in grades 6-8. In fact, being in the higher grade level category increased missingness by about 3 percentage points across all items.

Figure 1. Missingness, by grades 6-8 and 9-12.



Selected demographic variables were correlated with the number of missing items. Table 2.2 shows the correlation coefficients for these variables. Student ethnicity was dummy coded for this analysis. Most of the demographic variables showed modest or negligible correlations. Being Hispanic or a race other than White, African American, or Hispanic yielded virtually no relationship to missingness at all. (Because of the sample size, virtually all of the correlations were statistically significant at the $P < .05$ level or better; all P values should be interpreted with caution.) The three findings worth reporting are: 1) that grade level was positively correlated with missingness; 2) that being African American was positively correlated with missingness; and 3) that being male was positively correlated with missingness—yet, each of these correlations was small.

Table 2. Pearson correlations between demographic characteristics and amount of missing data

Demographic characteristic	Correlation with missingness
Sex (1 = Male, 2 = Female)	-0.07
Grade level	0.06
White	-0.12
Hispanic	0.00
African American	0.10
Other ethnicity	0.00

C. Creation of the Final Data Sets

To create the final data sets used for the factor analysis, survey forms in the top four percent of the missingness distribution were eliminated. This meant eliminating survey forms with 52 or more missing items (36% of the items). Using this filter, a total of 4% of the surveys were eliminated. Although eliminating 4% of the surveys was somewhat aggressive for a missingness threshold, the large data set allowed working with a nearly complete data set with minimal missing items.

D. The LES Scale Factor Analyses

Item and Scale Names

Survey item variable names, as reported in LES FactorAnalysisResults.pdf, are abbreviated to reflect the content of the information they seek. A trailing 'reversed' on the variable name indicates that it was reverse coded for consistency in interpretation. The scale names developed from the factor analyses were given names which simply describe the nature of the scale.

Conduct of the Factor Analyses

The factor analysis results are included in the file, LES FactorAnalysisResults 2011.pdf. Within the PDF file, factor analysis results are presented using the following format. Within the blue bordered box in the PDF, the following information is provided.

SCALE	The variable name of the scale developed during the SPSS runs.
ITEM POOL	A list of all of the items initially entered into the FA.
ITEMS IN THE ANALYSIS	A list of the items used for the specific factor analysis. Initially it was always the full item pool. Then, in subsequent analyses, selected subsets of items were employed.
N Fact.	The number of factors with eigenvalues > 1.0 that were identified in the analysis. The goal was to create scales with a single identified factor accounting for a large proportion of variance with good reliability.
Pct. Var.	The percentage of variance accounted for by the identified factor(s).
Alpha	Cronbach's alpha was the measure of reliability for these analyses.

A standard approach was applied to the factor analysis. Initially, all items from the item pool were included in the analysis. Sometimes only one factor emerged from the analysis. In that case, an attempt was made to eliminate items that shared minimal variance with the identified factor. The effect of the removal was an increase in the percent of variance accounted for by all of the items and usually an increase, or at least no real decrease, in Cronbach's alpha. At each

step, the items included in the analysis are shown. To the right of the blue box is a list of the final items included in the analysis.

If more than one factor was identified in the initial factor analysis, an attempt was made to arrive at a solution which included only a single meaningful factor. This was successfully achieved in all but three scales.

Overall, the majority of LES scales (17) had good reliability and account for substantial amounts of variance from their contributing items; these scales are listed in Table 3. Recommendations were made for modifying or eliminating scales which displayed low internal consistency (Cronbach's alpha) and/or had very few items representing the scale.

Table 3. LES scales with good reliability

- School climate
- Teacher and student respect
- Student discipline
- School safety
- Teacher to student relationships
- Teacher collaboration
- Student engagement
- Student encouragement
- Frequency of ATOD use
- Effect of ATOD
- Age of first use
- Perceived harmful effects of ATOD
- Parents' feelings toward ATOD
- Place of use
- Time of use
- Violence
- Bullying