WYOMING ALTERNATIVE SCHOOLS STUDENT CLIMATE SURVEY PILOT RESULTS

By
Michael Flicek, EdD
Consultant to WDE
Christopher Gwerder, MA
WDE Statistician

(January 3, 2017)

A Technical Advisory Group (TAG) was established by the Wyoming legislature in 2015. The work of the TAG was facilitated by staff from the National Center for Assessment (NCIEA). This group produced a report during the 2015 legislative interim and a second report during the 2016 interim. The 2016 report provided guidance on the development of a pilot accountability model to be implemented during the 2016-2017 academic year.

One part of that alternative school accountability pilot involves the administration of a student climate survey once in the fall and once in the spring. Ambitious participation rate goals were established for the survey. The goal was 95% participation for all students attending a school. The participation rate calculations will be based upon student enrollment data which will not be available until late spring or early summer. While 95% participation was likely not achieved statewide, preliminary information suggests that participation was quite high for a survey study. This paper provides suggestions for survey scoring and interpretation based upon analyses of the initial fall administration of the survey which was completed in October.

The survey was developed by the TAG using information from a variety of sources. Alternative schools serve many students with a history of low success in school and focus on the task of keeping these students engaged in and attending school. With this in mind, items were developed/selected for the survey that measured the relationship constructs of respect and support. There was one item that addressed trust which is an aspect of support. Other items were included on the survey to address academic rigor. To get at student perceptions of rigor, items were developed/selected that addressed high expectations. As such, the conceptual dimensions that guided the development of items for the survey were respect, support, and high expectations.

Scoring survey results may involve computing a total score on the survey. Often, however, subscale scores are computed as well. One way to organize items for subscales is to base them upon the apriori conceptual dimensions used in the survey's development. Another approach is to determine if empirical dimensions can be identified that are different in any important way from the conceptual dimensions.

This investigation used exploratory factor analysis to determine if empirical dimensions (i.e., subscales) could be identified based upon student responses to the survey. Confirmatory factor analyses were then performed to compare the model fit of the empirically identified subscales with the model fit of the subscales based upon the apriori established conceptual dimensions. An additional purpose of this investigation was to determine if school scores on the survey had enough variance for meaningfully differentiate among Wyoming alternative schools.

Method

Sample

To address these questions the responses of 600 alternative high school students from the first fall administration of the survey were used. Only students who responded to all 20 items were included. There were 6 students who indicated that their responses were not accurate reflections of their true views. These student's responses were excluded from the sample.

Results

Exploratory Factor Analysis

Principal components analysis was performed initially. The results of the principle components analysis are presented in Table 1. A scree plot that resulted from the principle components analysis is presented in Figure 1. Inspection of Table 1 and Figure 1 suggested a three factor solution fit the data well, given that it explains a majority of the variance and adding additional components yields little improvement.

Table 1. Results of the Principle Component Analysis.

Factors	Standard Deviation	Proportion of Variance	Cumulative Proportion
PC1	2.9887	0.4466	0.4466
PC2	1.2946	0.0838	0.5304
PC3	1.1574	0.0670	0.5974
PC4	0.8867	0.0393	0.6367
PC5	0.8683	0.0377	0.6744
PC6	0.8166	0.0333	0.7077
PC7	0.7827	0.0306	0.7384
PC8	0.7634	0.0291	0.7675
PC9	0.7166	0.0257	0.7932
PC10	0.7063	0.0249	0.8181
PC11	0.6853	0.0235	0.8416
PC12	0.6725	0.0226	0.8642
PC13	0.6645	0.0221	0.8863
PC14	0.6296	0.0198	0.9061
PC15	0.6119	0.0187	0.9248
PC16	0.5822	0.0170	0.9418
PC17	0.5727	0.0164	0.9582
PC18	0.5542	0.0154	0.9735
PC19	0.5419	0.0147	0.9882
PC20	0.4853	0.0118	1.0000

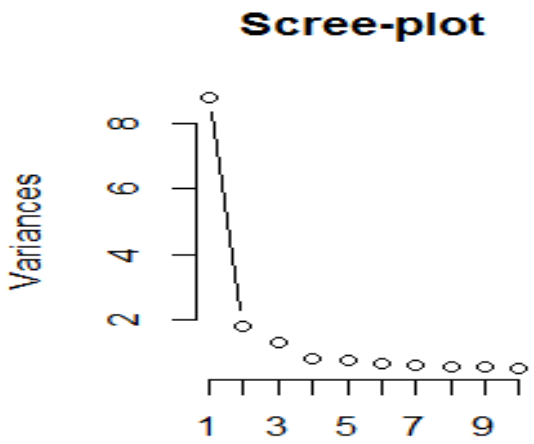

Figure 1. Scree Plot

Next, factor analysis, using an oblique rotation and the three factor solution was performed. These three factors explained 53% of the variance in student responses (see Table 2). The survey items, their conceptual dimensions and the factor loadings on the three empirical dimensions identified by this study are presented in Table 3.

Table 2. Variance Explained by Three Empirical Dimensions.

	Factor 1	Factor 2	Factor 3
SS loadings	5.83	2.55	2.13
Proportion Variance	0.29	0.13	0.11
Cumulative Variance	0.29	0.42	0.53
Proportion Explained	0.55	0.24	0.20
Cumulative Proportion	0.55	0.80	1.00

Inspection of Table 3 suggests that two of the empirically identified subscales differ from two of the conceptually identified subscales in some important ways. When items were developed it was assumed that the dimensions of respect and support within a school would be evident in both staff and student behaviors. As such, no distinction was made between whether the constructs of support and respect reflected staff versus student behaviors. Furthermore, it was assumed that respect and support would fall along different dimensions. The empirical dimensions, however, tell a different story. First, and perhaps most importantly, the empirical dimensions suggest that perceptions of staff behavior are on a separate dimension from perceptions of student behavior. As such, two of the empirical dimensions identified were labeled *Staff Respect/Support* and *Student Respect/Support.* The Staff Respect/Support subscale contains 12 items. All items on this subscale measure student perceptions of staff behaviors. There are 3 staff respect items, 8 staff support items and one staff trust item on this subscale. The finding that both support items and respect items loaded on the same factor suggests that student perceptions of these dimensions were strongly related. It is not particularly surprising that staff that are seen as respectful are also seen as supportive. It makes sense to place these items on the same subscale for the purpose of scoring and reporting. The Staff Respect/Support subscale accounted for 55% of the explained variance.

The Student Respect/Support subscale contained four items. All items on this subscale measure student perceptions of student behaviors. There are 3 student respect items and one student support item on this subscale. The Student Respect/Support subscale accounted for 24% of the explained variance. Finally, there were 4 rigor items on the conceptual dimension for *High Expectations* and these four items also loaded on an empirical dimension for High Expectations. This empirical dimension accounted for 20% of the explained variance. Thus, the empirical dimensions labeled here as (1) *Staff Respect/Support*, (2) *Student Respect/Support*, and (3) *High Expectations* differ in some important ways from the conceptual dimensions used for item development but they are, nevertheless, reasonable and interpretable item groupings for use as subscales on this survey.

Table 3. Factor Loadings for a Three Factor Solution.

Item Number	Text	Item Development Model	Empirical Model*		
			Factor 1 loadings (Staff Respect/Support)	Factor 2 loadings (Student Respect/Support)	Factor 3 loadings (High Expectations)
1	Teachers at this school believe I can perform well on challenging academic work.	Support	***0.41***	-0.03	0.22
3	I trust the staff at this school.	Support/Trust	***0.69***	0.11	-0.07
7	Teachers at this school do not let students give up when the work gets hard.	Support	***0.53***	0.19	0.07
8	There is at least one staff member at this school who knows me well and shows interest in my education and future.	Support	***0.59***	-0.14	0.11
9	Staff work hard to make sure that students stay in school.	Support	***0.55***	0.15	0.12
12	Students at this school are treated with respect by staff.	Respect	***0.85***	0.06	-0.12
13	Teachers give me helpful suggestions about how I can improve my work in class.	Support	***0.68***	0.00	0.10
16	Staff at this school treat me with respect	Respect	***0.88***	-0.06	-0.04
17	Staff at this school help students when they need it.	Support	***0.77***	0.03	0.06
18	Staff at this school make sure that I am planning for life after high school.	Support	***0.54***	0.01	0.20

19	Staff at this school treat each other with respect.	Respect	***0.60***	-0.01	0.17
20	Teachers explain things in a different way if students don't understand something.	Support	***0.73***	0.04	0.02
5	Students at this school help each other even if they are not friends.	Support	0.13	***0.61***	0.05
6	Students at this school treat property with respect.	Respect	0.04	***0.73***	0.03
10	Students at this school treat staff with respect.	Respect	-0.04	***0.78***	0.03
11	Students at this school treat each other with respect.	Respect	0.01	***0.84***	-0.03
2	Teachers at this school set high standards for academic performance.	High Expectations	-0.03	0.07	***0.62***
4	Students have to work hard to do well at this school.	High Expectations	-0.14	0.16	***0.53***
14	Teachers at this school expect students to do their best all of the time.	High Expectations	0.11	0.01	***0.64***
15	Teachers at this school have high expectations for me.	High Expectations	0.13	-0.04	***0.70***

*The factor loadings for each column that are bold and italicized indicate the corresponding item belongs on that empirical dimension.

Confirmatory Factor Analysis

Confirmatory factor analysis was performed by comparing absolute fit indices for four potential scoring and reporting models in order to demonstrate which potential model had the superior fit. One potential model was the factor structure identified by the exploratory factor analysis which is labeled as the *empirical model* in Table 3. A second potential model that was compared is labeled *the item development model* in Table 3. A third potential model was a two factor model. This model was identified by constraining the exploratory factor analysis to two factors. The resulting model had one factor for the four high expectation items and all other respect and support items were on the second factor. Finally, a one factor model where all items were on the one factor was included in for comparison.

Absolute fit indices were computed for each of the four potential scoring and reporting models. The results are presented in Table 4. First, a Chi-Square value was computed to assess "the magnitude of discrepancy between the sample and the fitted covariance matrices" (Hu & Bentler, 1999). Lower Chi-Square values represent better fit. The Chi-Square values for all four models were statistically significant at the $p < .001$ level. With a sample of 600 respondents this is not particularly surprising. From Table 4 we see that the empirical model had the lowest Chi-Square value.

Second, the comparative fit index (CFI: Bentler, 1990) was computed. The CFI is an index that is least affected by sample size. A CFI value of >= 0.95 is considered to represent good model fit (Hu and Bentler, 1999). From Table 4 we see that the CFI for the empirical model was 0.95 while the CFI values for the other three models were each less than 0.95.

Table 4. Model Fit Comparisons for Four Scoring and Reporting Structures

Statistic	Empirical Model	Item Development Model	2-Factor Model	1-Factor Model
Chi-Square	663	1,480	998	1,829
df	167	167	169	170
P-value	0	0	0	0
CFI	0.95	0.869	0.917	0.834
SRMR	0.039	0.064	0.05	0.069
RMSEA 90% CI	(.052, .061)	(.087, .096)	(.068, .077)	(.098, .106)
AIC	30,371	31,188	30,702	31,531

Third, a standardized root mean square residual (SRMR) was computed for each of the 4 models. While a SRMR of 0 indicates a perfect fit, well-fitting models generally have SRMRs of less than .05 (Byrne, 1998). Again, the empirical model has the lowest SRMR of 0.039 which was below 0.05. Fourth, root mean square error of approximation (RMSEA) confidence intervals were computed for each of the four models. Well-fitting models are those with lower bounds nearest to zero and higher bounds less than .08 (Hooper, Coughlan & Mullen, 2008). Both the empirical model and the 2-factor model have upper bounds less than .08 but the upper bound of the empirical model is lower than the lower bound of the 2-factor model. Finally, parsimony of fit was addressed with the Akaike Information Criterion (AIC). Generally the model with the lower AIC value has a better fit (Hooper et al., 2008). From Table 4 we see that the empirical model had the better fit.

Thus, on all absolute fit indices reported in Table 4 the empirical model with three factors had a better fit than any of the other models. These findings support recommending use of the three empirically identified factors for scoring and reporting results from the alternative school student survey.

Scores on the Empirical Model Subscales

Scores for each student were computed on each of the three subscales for the empirical model. Items were scored as 1 for “strongly disagree”, 2 for “disagree”, 3 for “agree” and 4 for “strongly agree”. The student scores were the mean of the subscale’s item scores. Figure 2 provides boxplots for the frequency distributions on the three subscales for the 600 alternative school students who responded to all items on the survey. Table 5 presents the Pearson correlation coefficients for the student scores on the 3 empirical dimensions. The coefficients in Table 5 are about what would be expected in that they show the subscales are related to a larger school climate construct but are somewhat unique as well.

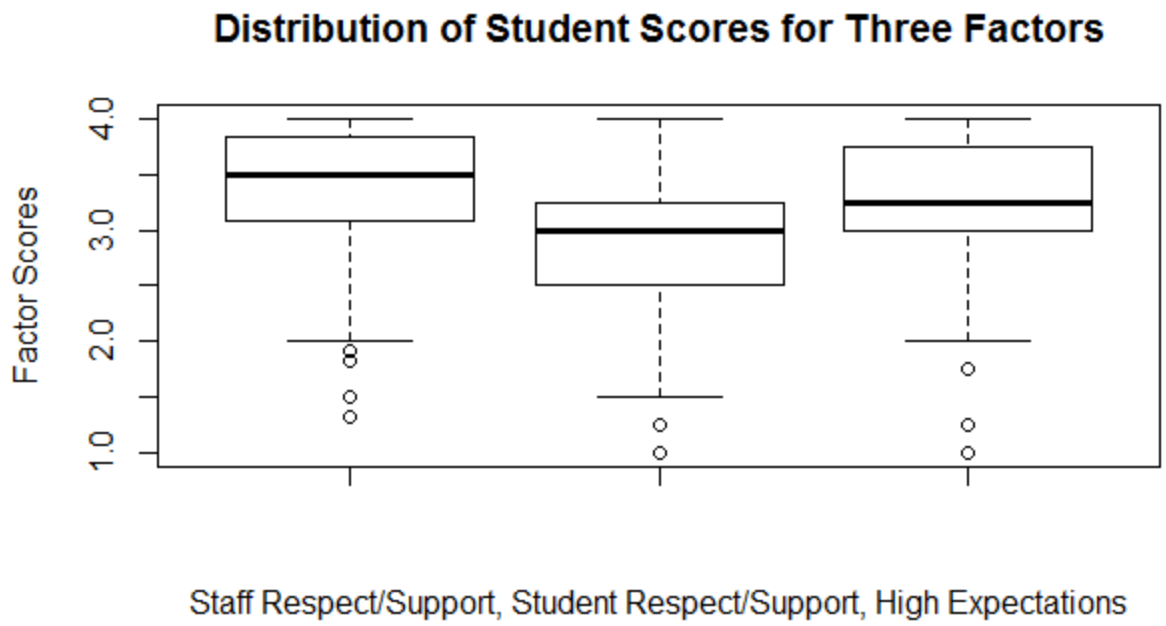

Figure 2

Table 5. Pearson Correlation Coefficients for the Empirical Dimensions.

	Student Respect/Support	High Expectations
Staff Respect/Support	0.54	0.65
Student Respect/Support		0.36

Evidence for Meaningful Differentiation

A prominent feature in the federal every student succeeds act (ESSA) is the stipulation that school accountability indicators must be able to meaningfully differentiate among schools. If all schools being measured do universally well or universally poorly on an indicator, that particular indicator does not add much value to the school accountability model. The ability of school scores, on the three empirical subscales from this survey, to meaningfully differentiate among the Wyoming alternative schools was therefore of interest.

The school scores studied here were the mean scores of all students attending the school on each of the 3 subscales. Mean scores and standard deviations were also computed for the entire sample of alternative high school students in Wyoming. Cohen's *d* effect sizes were also computed by subtracting the school mean from the state mean and dividing by the state standard deviation. These effect sizes indicate how far the school's mean is above or below the state mean expressed as a percentage of the state standard deviation. The state alternative high school mean and standard deviation are presented in Table 6.

Table 6. Statewide Mean and Standard Deviation for Sample of Wyoming Students Who Completed the Student Climate Survey in the October, 2016.

	Mean	Standard Deviation
Staff Support/Respect	3.32	0.49
Student Respect/Support	2.80	0.67
High Expectations	3.23	0.46

Collectively the mean score on Staff Support/Respect was the highest and the Student Respect/Support mean score was the lowest. The mean for High Expectations fell between the other two means but it was closer to Staff Support/Respect than to Student Respect/Support. These findings are reasonable in that both Staff Support/Respect and High Expectations involve perceptions of teacher actions and Student Respect/Support reflect perceptions of student actions.

The school means and effect sizes on each of the three empirical dimensions are presented in Table 7. Five of the alternative schools had fewer than 10 respondents to the survey. For the purpose of this analysis, those schools were excluded. From Table 7 we see that the school effect sizes on Staff Support/Respect ranged from 54% of a standard deviation below the state mean to 82% of a standard deviation above the state mean for a total range from the lowest score to the highest score of 1.36 standard deviation units. The school effect sizes on Student Respect/Support ranged from 148% of a standard deviation below the state mean to 111% of a standard deviation above the state mean for a total range of 2.59 standard deviation units. On High Expectations the means ranged from 62% of a standard deviation below the state mean to 66% of a standard deviation above the mean for a total range of 1.28 standard deviation units. There is ample variance on these dimensions to meaningfully differentiate among the Wyoming alternative schools.

Table 7. School Means and Effect Sizes for the Three Empirical Dimensions.

School	Respondent Count	Staff Support/Respect		Student Respect/Support		High Expectations	
		Mean	Effect Size	Mean	Effect Size	Mean	Effect Size
One	18	3.66	0.69	3.15	0.53	3.36	0.29
Two	73	3.49	0.35	3.01	0.31	3.43	0.44
Three	30	3.57	0.50	3.01	0.31	3.53	0.66
Four	32	3.31	-0.02	2.72	-0.12	2.95	-0.62
Five	35	3.63	0.63	3.06	0.39	3.36	0.29
Six	22	3.73	0.84	3.55	1.11	3.50	0.59
Seven	17	3.49	0.34	2.71	-0.14	3.46	0.49
Eight	63	3.46	0.28	2.96	0.24	3.35	0.27
Nine	17	3.36	0.08	2.62	-0.27	3.34	0.24
Ten	18	3.71	0.79	2.90	0.15	3.21	-0.05
Eleven	81	3.54	0.45	3.06	0.39	3.39	0.35
Twelve	69	3.13	-0.38	2.50	-0.45	3.16	-0.16
Thirteen	17	3.40	0.16	3.07	0.41	3.24	0.01
Fourteen	13	3.06	-0.54	1.81	-1.48	3.00	-0.50
Fifteen	39	3.43	0.23	2.73	-0.10	3.28	0.11
Sixteen	25	3.27	-0.10	2.47	-0.49	3.12	-0.24

Conclusions and Recommendations

Three empirical dimensions were identified that were substantively different from the conceptual dimensions that guided item development. First, items related to staff support and staff respect all loaded on one single dimension. This empirical Staff Respect/Support dimension has 12 items while the other two dimensions had 4 items each. More items typically result in a more reliable score and the Staff Respect/Support subscale is likely the most reliable of the three dimensions. Having a supportive and respectful staff is particularly important for alternative schools that are largely serving students with a history of low success in schools. An important role for alternative schools is to keep these students engaged in school and help them graduate. Students perceiving the staff to be supportive and respectful would make an important contribution to that goal.

Second, separating the perceptions of student from teacher actions, related to respect and support, is a substantial but important change given the fact that student perceptions were quite different on these two dimensions. Both dimensions are actionable and gains on one or both would be compelling evidence of an improving climate in a school. More supportive and respectful students and staff improve the school's ability to focus on the student learning mission.

Third, the conceptual dimension of High Expectations was also identified as a distinct empirical dimension. The support and respect dimensions set the stage for learning to occur and the high expectations dimension provides evidence of rigor at a school. Taken together these three empirical dimensions provide important information about a school's climate that can inform school improvement decisions.

Finally, there was evidence that school scores on these dimensions had ample variance for the purpose of meaningfully differentiating among the alternative schools in Wyoming.

In light of these findings we recommend that these three empirical subscales be used for scoring and reporting of survey results both for this pilot year and, in the absence of data suggesting otherwise, future operational administrations.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107 (2),* 238-246.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS* and *SIMPLIS: Basic concepts, applications and programming.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hooper, D., Coughlan, J. and Mullen, M. R. (2008). Structural equation modeling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6 (1),* 53-60.

Hu, L. T. and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6 (1),* 1-55.