Abstract Title Page – Paper 2

Title: Assessing the Psychometric Characteristics of a Child/Adolescent Behavioral and Emotional Risk Self-Report Screener Across Grade Levels and School Districts

Authors and Affiliations:
Harrell-Williams, L. M., Georgia State University
Dowdy, E., University of California - Santa Barbara
Twyford, J., California Lutheran University

Abstract Body

Background / Context: The BESS is a relatively new screening system for identifying behavior and emotional risk (BER) in children and adolescents. Psychometric evidence regarding this instrument is important for researchers and practitioners considering the use of the BESS for identifying BER in students. This study seeks to provide and evaluate evidence for the use of the Student form of the BESS.

Purpose / Objective / Research Question / Focus of Study
The objectives of this study are to:
(1) Assess the internal structure of the BESS Student Form using factor analysis
(2) Determine if the items perform differently across two locations with different ethnic characteristics via Rasch-based DIF analysis
(3) Assess the odds of elevated risk levels based on student demographics

Settings
Data were collected in two locations: the Los Angeles United School District (LAUSD) and Bibb County (GA). LAUSD serves the largest school population of the public school systems in California (approximately 700,000 students). It is the second largest school district in the United States, after the New York City Department of Education. Overall, current school statistics report that 74% of the student population is Hispanic, 11% African-American, 9% White, and 4% Asian. Over half of the student population is considered economically disadvantaged and a large proportion of the total student population is classified as Limited in English Proficiency. The Bibb County (GA) School system serves the residents of Macon and Bibb County, with an enrollment of approximately 25,000 students across PreK-12. Approximately 75% of the students identify as African American/Black, 22% identify as White/Caucasian (non-Hispanic), and 3% identify as another race/ethnicity. Approximately 77% of students in the district qualify for free/reduced lunch benefits.

Population / Participants / Subjects
The first sample was collected from the Los Angeles United School District (LAUSD) as part of a larger IES funded research project investigating the longitudinal validity for screening (Advancing Children's and Teacher's Success through Early Screening and Intervention; U.S. Department of Education, R324B060005). The sample of 273 elementary and middle school students (grades 3 through 7, and ages 7–12) had approximately 52% male participants. Ethnicity was reported by parents as follows: 81.4% Hispanic, 4.5% White, 2.6% African American, 1.9% Filipino, and 1.5% Asian.

The second sample contained the responses of 4,074 students from 3 LAUSD high schools in the Northern region of the school district. Approximately 54% of the students were male. Approximately 26% of the students were 9th graders, 26% were 10th graders, 27% were 11th graders and 21% were 12th graders. School report data regarding variables such as ethnicity are not available at this time but the percentages should be roughly equivalent to that seen in the elementary and middle school students.

The third sample was collected as part of a pilot for district-wide screening initiative in Bibb County. There were 1,874 students across the four schools in grades 9 through 12. There were approximately 47% males in the sample. Ethnicity was obtained from school records: 72.8% African American, 24% White, 1.6% Hispanic, 0.4% Asian, with the remaining 1.1% identified as "Other".

Intervention / Program / Practice

The BASC-2 Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007) measures behavioral and emotional strengths and weakness in children and adolescents. The BESS Student form is a 30 item self-report form, and according to readability indexes, is appropriate for students in grades 3 through 12. The items ask the student to rate how frequently they think or feel a certain way using a 4 point rating scale (Never, Sometimes, Often, Almost Always). The sum of the item raw scores is transformed into a linear T score, which is used to categorize the students as exhibiting normal risk, elevated risk or extremely elevated behavioral and emotional risk.

Research Design

To evaluate the constructs being measured by the BESS as well as evaluate the predictive validity for future educational outcomes, a longitudinal, trend survey approach was used. Data was collected across multiple grades for three years for children in the Third through Seventh grades, but the participants in each year did not remain the same. To evaluate the appropriateness of the BESS in high school-aged students, a cross-sectional survey approach was used. Data was collected at one time point for all high schools.

Data Collection and Analysis

Data Collection:

Sample 1: During three school years covering the 2008 – 2010 timeframe, the research team administered the BESS Student self-report form to select elementary and middle school classrooms in LAUSD. The number of schools varied by year, with 17 schools participating in the first year, 27 schools during the second year, and 22 schools during the third year. Four students (2 males and 2 females) were randomly selected from each classroom that participated. Research team members, including IES-funded researchers and one IES funded post-doctoral researcher (IES Grant: R324B080006) used a prescribed protocol and script with each classroom group. This approach resulted in 2,829 participants over the span of the longitudinal study.

Sample 2: The BESS Student was administered in every classroom at the 3 participating high schools in September, 2012. Research team members, including two IES funded post-doctoral researchers (IES Grant: R324B080006), and graduate students from the University of Santa Barbara used a prescribed protocol and script with each classroom group. This resulted in 4,074 participants.

Sample 3: The BESS Student was administered in every classroom at the 4 participating high schools during the 2009 – 2010 school year. Research team members, including an IES funded post-doctoral researcher (IES Grant: R324B080006), used a prescribed protocol and script with each classroom group.

Analysis:

Factor and Item analysis (Research Question 1): The data from the first year of the longitudinal study (Sample 1) was used to perform exploratory and confirmatory factor analysis (Dowdy et al., 2011) using SEM-based procedures in MPLUS. The LAUSD and Bibb high school data were used to perform a confirmatory factor analysis of the theorized four factors of the BESS, as well as a Rasch-based item analysis. Conquest (Wu, Adams, Wilson, & Heldane, 2007) was used to fit a unidimensional and a four-dimensional structure using a multidimensional partial credit model. Lower values of AIC (Akaike Information Criteria, Akaike, 1973; Akaike, 1987) and BIC (Bayesian Information Criteria; Schwarz, 1978) values served as indicators of model fit. Ideally, item mean-square fit indices should be between .5 and 1.5 (Linacre, 2004).
Differential Item Functioning Analysis (Research Question 2): Since the two school districts have markedly different ethnicity demographics, a Rasch-based differential item functioning (DIF) analysis for groups via IRTPRO software (Cai, du Toit, & Thissen, 2011) was used to evaluate measurement invariance across the two school districts.
***Logistic Regression Analysis (Research Question 3)*:** Data from all years of the longitudinal study as well as the Bibb data were used in the logistic regression analysis to estimate the odds of elevated risk by demographics (gender, grade, ethnicity, and free lunch status). Odds ratios were used as effect size indices, with the cut-off for large effect sizes being defined as an odds ratio greater than 1.89 (or less than .53, for negative relationships) and moderate effect sizes being an odds ratio greater than 1.53 (or less than .65, for negative relationships) (Monahan, McHorney, Stump, and Perkins, 2007).

Findings / Results

Research Question 1 – Factor and Item Analysis

Sample from Elementary/Middle School Dataset

Dowdy et al. (2011) presented a simple-structure four factor solution, with items loading on Personal Adjustment, Inattention/Hyperactivity, Internalizing Problems, and School Problems factors (see Figure 1). Item 9 (being liked by others), Item 11 (difficulty sitting still), and Item 22 (feeling stupid) were removed during the analysis due to inadmissible factor loadings greater than 1.0, and/or factor cross-loading both positively and negatively. Goodness-of-fit statistics supported the four factor model with the set of 27 items, c^2 (249) = 528.705, $p < .00$, RMSEA = .038 (90% confidence interval [CI]: .33–.042), and SRMR = .028.

(Insert Figure 1 here)

Samples from the High School Data

As shown in Table 1 below, AIC and BIC both provide evidence for the four factor structure (Figure 2) over the unidimensional structure in analyses using the LAUSD and the Bibb data. Analyses were conducted using all 30 items on the BESS Student form. Three of the factors have EAP/PV reliabilities that are below .80, which is lower than generally accepted values for a score inference of this importance; the presence or absence of BER for development of a mental health disorder that may, in turn, adversely affect educational outcomes. The item analysis started with the output for all 30 BESS items. All 30 items had weighted and unweighted mean square fit statistics in between the limits listed as productive for measurement (.5 to 1.5). Based on these values, the removal of items 9, 11, and 22, as in Dowdy et al. (2011), did not seem necessary. Based on these results, all 30 items are used for the remaining analyses. Given that three of the

four factors had reliabilities below .80 (Table 2), the current central score inference of reporting behavioral and emotional risk using one overall norm referenced T score is still appropriate.

(Insert Tables 1 & 2, and Figure 2 here)

Research Question 2: *Rasch-based DIF analysis for Comparison of School Districts*
As the demographics of the students differ greatly between the districts, a Rasch-based DIF analysis was conducted using IRTPRO. The test for each item revealed that DIF was not present for any of the 30 items across the two school districts for the high school data, as noted by the non-significant p-values presented in Table 3.

(Insert Table 3 here)

Research Question 3
Logistic Regression Analysis of Risk and Demographics
The results of the regression analysis are shown in Table 4. While gender, grade level, and race were significant predictors based on Wald tests, only the "Whites to Others (Other than Non-African American/Hispanic)" had a notable effect size using the criteria specified in Monahan et al. (2007). The odds ratio of 2.629 means that white students were more likely than those classified as "Other" (which included Asians, Filipino, American Indian/Alaskan Native, and Pacific Islander) to have elevated risk.

(Insert Table 4 here)

Conclusions
First, the theorized four factor structure of the BESS holds for all school levels (elementary/middle and high school). Unlike Dowdy et al (2011), all 30 items are retained in the factor structure as the weighted and unweighted mean square fit statistics were reasonable. However, in the analysis of the high school data, three of the four factors had reliabilities below .80, indicating that the current method of reporting behavioral and emotional risk using one overall T score is still appropriate.

Second, no evidence of differential item functioning was found across the 30 items when comparing the LAUSD high school group to the Bibb high school group. The items performed equivalently across the diverse populations measured in these two samples, which speaks to the ability to use the instrument across school districts with varying ethnic demographics.

Third, logistic regression analysis revealed that gender and location did not affect risk. Odds ratios showed that differences in risk due to grade level and most races were not practically significant. The large odds ratio noted in the "Whites vs. Others (Non-African American/Hispanic)" comparison leads to a need for larger sample sizes for those sub-populations sampled. In the longitudinal LAUSD and the Bibb samples used for this analysis, for example, whites were in the minority, with "Others" barely being represented.

Appendices

Appendix A. References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), Second international symposium on information theory (pp. 267–281). Budapest: Akademiai Kiado.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317–332.

Cai, L., du Toit, S. H. C., & Thissen, D. (2011). IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling [Computer software]. Chicago: Scientific Software International.

Dowdy, E., Twyford, J. M., Chin, J. K., DiStrefano, C. A., Kamphaus, R. W., and Mays, K.L. (2011). Factor Structure of the BASC–2 Behavioral and Emotional Screening System Student Form. *Psychological Assessment, 23*(2). 379 – 387.

Kamphaus, R. W., & Reynolds, C. R. (2007). Behavior Assessment System for Children—Second Edition (BASC–2): Behavioral and Emotional Screening System (BESS). Bloomington, MN: Pearson.

Linacre, J. M. (2004). Optimal rating scale category effectiveness. In E. V. Smith, Jr. & R. M. Smith (Eds.), Introduction to Rasch measurement (pp. 258-278). Maple Grove, MN: JAM Press.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6(2),* 461-464.

Wu, M. L., Adams, R. J., Wilson, M. R., & Heldane, S.A. (2007). ACER ConQuest: Generalized item response modeling software (Version 2.0) [computer software]. Melbourne: Australian Council for Educational Research.

Appendix B. Tables and Figures

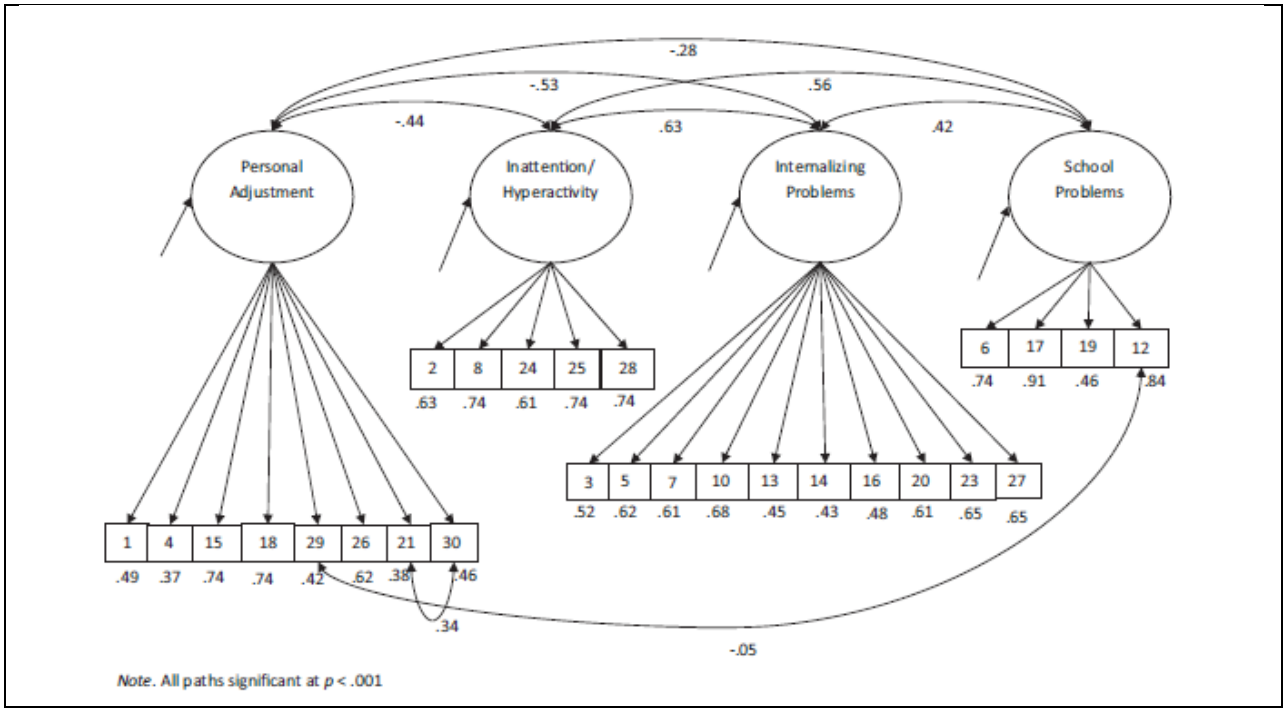

Figure 1. Confirmatory Factor Analysis Model, Dowdy et al. (2011)

Table 1. AIC and BIC Values for CFA Model Comparisons

Sample	Sample Size	Number of Factors	Deviance	Number of Parameters	AIC	BIC
LAUSD	4074	One	241262.1	91	241444.1	242018.5
		Four	235664.1	100	**235864.1**	**236495.3**
Bibb	1874	One	109403.8	91	109585.8	110160.2
		Four	106977.5	100	**107177.5**	**107808.8**

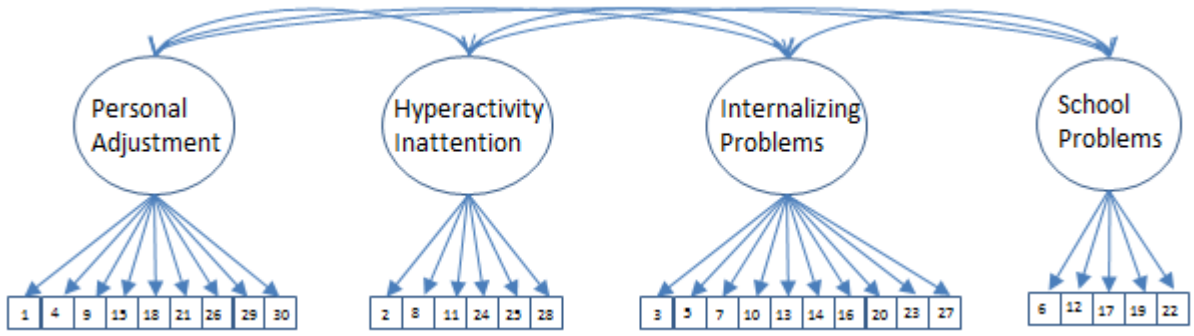

Figure 2. Confirmatory Factor Analysis Model, LAUSD High School Data

Table 2. EAP Reliabilities for Four Factor Structure

Sample	Personal Adjustment	Hyperactivity/ Inattention	Internalizing Problems	School Problems
LAUSD	.791	.728	.842	.758
Bibb	.788	.736	.832	.731

Table 3. DIF Testing Output

Group 1	Group 2	Total X^2	*d.f.*	*p*	X^2_a	*d.f.*	*p*	$X^2_{c\|a}$	*d.f.*	*p*
1	1	0.6	3	0.8960	-----	-----	-----	0.6	3	0.8960
2	2	0.1	3	0.9916	-----	-----	-----	0.1	3	0.9916
3	3	0.0	3	0.9988	-----	-----	-----	0.0	3	0.9988
4	4	2.4	3	0.5005	-----	-----	-----	2.4	3	0.5005
5	5	0.1	3	0.9939	-----	-----	-----	0.1	3	0.9939
6	6	0.0	3	0.9999	-----	-----	-----	0.0	3	0.9999
7	7	0.0	3	0.9981	-----	-----	-----	0.0	3	0.9981
8	8	0.0	3	0.9994	-----	-----	-----	0.0	3	0.9994
9	9	0.5	3	0.9116	-----	-----	-----	0.5	3	0.9116
10	10	0.1	3	0.9927	-----	-----	-----	0.1	3	0.9927
11	11	0.4	3	0.9459	-----	-----	-----	0.4	3	0.9459
12	12	0.2	3	0.9833	-----	-----	-----	0.2	3	0.9833
13	13	0.1	3	0.9917	-----	-----	-----	0.1	3	0.9917
14	14	0.1	3	0.9935	-----	-----	-----	0.1	3	0.9935
15	15	0.0	3	0.9986	-----	-----	-----	0.0	3	0.9986
16	16	0.0	3	0.9999	-----	-----	-----	0.0	3	0.9999
17	17	0.0	3	0.9995	-----	-----	-----	0.0	3	0.9995
18	18	0.0	3	0.9995	-----	-----	-----	0.0	3	0.9995
19	19	0.6	3	0.9005	-----	-----	-----	0.6	3	0.9005
20	20	0.2	3	0.9700	-----	-----	-----	0.2	3	0.9700
21	21	0.4	3	0.9446	-----	-----	-----	0.4	3	0.9446
22	22	0.1	3	0.9868	-----	-----	-----	0.1	3	0.9868
23	23	0.2	3	0.9792	-----	-----	-----	0.2	3	0.9792
24	24	0.1	3	0.9957	-----	-----	-----	0.1	3	0.9957
25	25	0.0	3	0.9991	-----	-----	-----	0.0	3	0.9991
26	26	0.4	3	0.9415	-----	-----	-----	0.4	3	0.9415
27	27	0.3	3	0.9640	-----	-----	-----	0.3	3	0.9640
28	28	0.5	3	0.9114	-----	-----	-----	0.5	3	0.9114
29	29	0.6	3	0.8964	-----	-----	-----	0.6	3	0.8964
30	30	0.1	3	0.9917	-----	-----	-----	0.1	3	0.9917

Table 4. Logistic Regression Output

Variable	Wald Test Statistic	df	p-value	Odds Ratio
Grade	8.861	1	.003	.904
Gender	2.318	1	.128	.877
Free Lunch Status	6.287	2	.043	
Paid vs. Free				.654
Reduced vs. Free				.791
Race	38.75	3	.000	
Black vs. Others				1.065
Hispanic vs. Others				1.004
White vs. Others				2.629
Location (LAUSD vs. Bibb)	-0.134	1	.664	.874