

Examining the Relationship Between Scores on the *Behavioral and Emotional Screening System* and Student Academic, Behavioral, and Engagement Outcomes: An Investigation of Concurrent Validity in Elementary School

Tyler L. Renshaw, Katie Eklund, Erin Dowdy, Shane R. Jimerson,
Shelley R. Hart, James Earhart, Jr., and Camille N. Jones
University of California, Santa Barbara

Universal screening of emotional and behavioral problems among students warrants further consideration by school professionals. School-based universal screening may provide opportunities for early identification and intervention, ultimately preventing the development of more severe problems and promoting more positive outcomes in the future. The *Behavioral and Emotional Screening System* (BESS) is a contemporary screening instrument that may be used to identify risk for emotional and behavioral problems in students from preschool to high school. The purpose of the present study was to examine the concurrent validity of the BESS in elementary school settings. Specifically, this study examined the relation between BESS ratings and report-card outcomes (i.e., academic, behavioral, and engagement marks). The results supported the hypotheses that students' risk-level classifications were significantly related to school-based outcome criteria and that school-based outcome criteria were deemed to be effective discriminators of students' risk-level classification. Limitations, future directions for research, and implications for practice are discussed herein.

Universal screening for students' emotional and behavioral problems is becoming an increasingly important activity for school systems to consider. Given that students with emotional and behavioral problems have poor school-related outcomes (Rones & Hoagwood, 2000), school-based screening may provide opportunities for early identification and intervention, ultimately preventing the development of more severe problems and promoting more positive outcomes in the future (Dowdy, Furlong, Eklund, Saeki, & Ritchey, in press). However, despite its ameliorative potential, only about 2% of schools in the United States implement universal screening efforts (Romer & McIntosh, 2005). Considering that schools often function as the de facto mental health care system for students and adolescents (Rones & Hoagwood), the school context affords a unique opportunity to systematically identify and provide support services for students with emotional and behavioral problems.

Screening and School Psychology

As data-based advocates for students, school psychologists can help identify students with emotional and behavioral risk by advocating for and implementing universal screening within their local schools. As they embark on such efforts, school psychologists should be cognizant of several key considerations (Dowdy et al., in press). First, universal screening should never be isolated – it should always be integrated within a larger student-support framework. Second, screening efforts should always be accompanied by well-defined objectives, including progress monitoring and service provision aims. Third, the pragmatics of screening implementation – who, when, and where – must be established through careful consideration and planning. And lastly, decisions must be made regarding which types of emotional and behavioral problems to screen for and, by extension, what screening instrument to use.

To date, there are several research-based instruments for school psychologists to utilize, though

not all are appropriate for every school context. Thus, when selecting a screening instrument, it is recommended that school psychologists first consider three key aspects (Glover & Albers, 2007): (a) the match between the screener, the objectives underlying screening, and the support system surrounding the screening process; (b) the technical adequacy – including sensitivity, specificity, positive and negative predictive values, and the psychometric properties of the instrument; and (c) the social validity (i.e., practicality and feasibility) of using the screener and managing the screening process amidst typical school duties and circumstances (e.g., Caldarella, Young, Richardson, Young, & Young, 2008). Mindful of the aforementioned considerations, as well as the nature of school psychological services, Dowdy et al. (in press) recommend four instruments as potentially useful for school-based universal screening: the *Strengths and Difficulties Questionnaire* (Goodman, 1997), *Pediatric Symptom Checklist* (Little, Murphy, Jellinek, Bishop, & Arnett, 1994), *Systematic Screening for Behavior Disorders* (Walker & Sevenson, 1992), and the *Behavior Assessment System for Children-2 Behavioral and Emotional Screening System* (BASC-2 BESS; Kamphaus & Reynolds, 2007).

BASC-2 Behavioral and Emotional Screening System

The BESS (Kamphaus & Reynolds, 2007) is the most recent and least-researched of the four recommended screening instruments. It is used to identify emotional and behavioral strengths and weaknesses in students from preschool to high school, assessing both internalizing and externalizing problems as well as school-related difficulties and adaptive skills. It has three parallel report forms – student, parent, and teacher – each composed of 25-30 items, designed to be completed in 5 minutes or less. The majority of items comprising the BESS stem from the item pool created during the development of the BASC-2 *Teacher Rating Scales*, *Parent Rating Scales*, and *Self-Report of Personality* (Reynolds & Kamphaus, 2004), with a range of one-to-eight new items added to each form. Similar to the other BASC-2 assessments, there are four response options for each item (i.e., never, sometimes, often, almost always); but dissimilar from the other assessments, the BESS only produces a single score. This score is conceptualized as the student's risk-level classification for emotional and behavioral problems and can fall within the range of one of three categories: *normal*, *elevated*, or *extremely elevated*. Although the BESS has yet to garner substantial empirical support, its characteristics suggest it will be a promising tool for school psychological practice. However, to ensure its effectiveness and social validity, investigations must first demonstrate its school-based, criterion-related validity – requiring concurrent and predictive evidence.

To date, only two published studies have examined the criterion-related validity of the BESS in school settings. The first, conducted by Kamphaus et al. (2007), used 2-year longitudinal data from students in K-5 grades to evaluate the predictive validity of the screener against a variety of behavioral and educational outcomes. Overall, correlations indicated that the screener was particularly good at predicting future teacher ratings of conduct problems, atypicality, and social skills; future indices of school maladjustment, special education placement, and referral for prereferral intervention; and reading and mathematics grades and standardized test scores. The second study, conducted by DiStefano and Kamphaus (2007), used longitudinal data from preschool students to evaluate the concurrent and predictive validity of the screener with various diagnostic and educational outcomes in kindergarten. Overall, correlations indicated that BESS scores were significantly related to concurrent assessments of students' behavioral symptoms and school readiness as well as predictive assessments of students' disciplinary infractions; grades for reading, social development, and work habits; BASC-2 teacher-reported subscales for externalizing, internalizing, adaptability, school problems, and behavioral symptoms; and standardized testing scores for reading and math. Given the limited nature of this research to date, additional studies are warranted to further examine the validity of the BESS.

Present Study

The purpose of the present study was to further evaluate the concurrent validity of the BESS in elementary school settings. Specifically, the intent was to examine the relation between teachers' BESS ratings with students' recent report-card outcomes. Overall, there were three hypotheses in this study:

1. Students' risk-level classifications – derived from screening results and grouped herein as either “normal” or “at-risk” – could be significantly correlated with their academic, engagement, and behavioral outcomes, as graded by their teachers.
2. There would be significant mean differences between “normal” and “at-risk” students' academic, behavioral, and engagement outcomes, showing the relevance of the BESS to school-based indicators.
3. Academic, engagement, and behavioral report-card criteria would effectively discriminate between students identified as “normal” and “at-risk” via screening results.

METHODS

Participants

Participants were 26 third-graders and 22 fourth-graders from two elementary schools in a suburban community, within the same school district, located on California's central coast. During the 2008-2009 school year, the total enrollment of one school was 286 students and the total enrollment of the other was 421 students. During that time, the demographic make up of both schools was comparable, with approximately 73% of students identifying as Hispanic or Latino, 18% as White, and 9% as other or multiple ethnic groups. Approximately 68% of the students were classified as socioeconomically disadvantaged, 40% as English language learners, and 14% as students with disabilities. Using class-wide data collection procedures, the demographics of the participants in the present study ($N = 48$) were representative of the student population in these schools.

Measures

BESS teacher form. The BESS teacher form (child/adolescent version) is completed by teachers of students in grades K-12 (Kamphaus & Reynolds, 2007). It consists of 25 items and is designed to be completed in 5 minutes or less per student. The screener is scored by summing the items to generate a total *T*-score, with lower scores (20-60) reflecting a “normal” level of risk, higher scores (61-70) reflecting an “elevated” level of risk, and still higher scores (71 or above) reflecting an “extremely elevated” level of risk. The BESS teacher form was developed and normed with a sample of 12,350 accompanying parent and student forms, derived from participants in 233 cities across 40 states. Results from the norming process indicate that the psychometric properties of the BESS (across all forms) are generally acceptable, having good split-half reliability (.90-.96), test-retest reliability (.80-.91), inter-rater reliability (.71-.83), sensitivity (.44-.82), and specificity (.90-.97). Furthermore, the measure has also proven to have acceptable convergent validity with the Achenbach System of Empirically Based Assessment (.71-.77), Conner's Rating Scales (.51-.78), Vineland Adaptive Behavior Scales (.32-.69), Children's Depression Inventory (.51), and the Revised Children's Manifest Anxiety Scale (.55).

Report cards. Students' report cards consisted of academic, engagement, and behavioral indicators, graded by their teachers. The academic indicators comprised 6 total subject areas – listening, reading, writing, math, history, and science – and corresponded to California state educational standards. Each indicator was graded on a scale of 1 to 4 (1 = has difficulty with standard, 2 = approaches standard, 3 = meets and applies standard, 4 = exceeds standard), indicating teachers' perceptions of students' present levels of achievement. For the purposes of this study, each subject area was conceptualized as a sub-composite, making up a total *Academic Achievement* composite. A behavioral engagement indicator accompanied each subject area, wherein the teachers graded the amount of “effort” students exhibited in meeting academic standards, using the same grading scale. Because these engagement indicators were unidimensional and few in number, for the purposes of this study they were summed into a total *Engagement* composite. The report card also consisted of several behavioral indicators (e.g., “Follows rules and

direction;” “Completes classwork;” “Works well in a group”), graded on a 1-to-3 scale (1 = needs improvement, 2 = satisfactory, 3 = excellent). Using the same rationale as the engagement indicators, these behavioral indicators were summed into a total *Behavioral Performance* composite.

Procedures

During the first quarter of the school year, the BESS teacher form was completed for all the third-graders attending one school and all the fourth-graders attending the other school. For both grades combined, screening outcomes indicated that 70% of students were in the *normal* range ($n = 77$), 18% were in the *elevated* range ($n = 20$), and 12% were in the *extremely elevated* range ($n = 13$). Thus, for the purposes of this study, students in the elevated and extremely elevated ranges were grouped together, resulting in dichotomized risk-level classification: *normal* (T -score of 20 to 60) or *at-risk* (T -score of 61 and above). Using this classification method, screening results indicated that 20 third-graders and 13 fourth-graders had BESS scores in the at-risk range. In an attempt to create matched groups, the 13 at-risk fourth-graders were selected to participate in the study, matched with a random selection of 13 normal fourth-graders. A random selection of 13 at-risk third-graders was then conducted, matched with a random selection of 13 normal third-graders. During the course of the study, however, 2 at-risk fourth-graders were transferred to another school, and so the matched pairs were reduced to 11 fourth-graders and 13 third-graders in each group ($N = 48$).

Next, the sample participants’ first quarter report cards – graded within a few weeks of BESS completion – were examined and coded. The Listening, Reading, Writing, Math, History, and Science sub-composites were generated and weighted by summing the indicators associated with each subject area and then dividing that total by the respective number of indicators. The Academic Achievement, Engagement, and Behavioral Performance composites were derived via the same process as the sub-composites, using their respective indicators. Following data collection and preparation, the aforementioned hypotheses were examined by conducting three sets of statistical analyses: bivariate correlations, a one-way ANOVA, and discriminant function analyses. All analyses utilized the previously described sub-composites or the general composites; no isolated indicators were included in the analyses.

RESULTS

Results indicated that teacher-rated BESS risk-level classification (i.e., either *normal* or *at-risk*) was significantly related with students’ concurrent academic, engagement, and behavioral outcomes, as reported on their report cards. Specifically, risk-level classification was significantly correlated with each academic sub-composite, contributing to a significant correlation with the overall Academic Achievement composite. In addition, significant correlations were also found between risk level and the Engagement and Behavioral Performance composites. Furthermore, results from the one-way ANOVA indicated significant differences between the mean scores of the normal and at-risk students for each of the sub-composites as well as the overall Academic Achievement, Engagement, and Behavioral Performance composites. See Table 1 and Table 2 for statistical summaries of these results.

Two separate discriminant function analyses were conducted to determine if the academic, engagement, and behavioral composites, derived from the recent report cards, were effective criteria for discriminating between students’ risk-level classification. *Model 1* was theoretically driven, entering the Academic Achievement, Engagement, and Behavioral Performance composites as simultaneous discriminant criteria. Results revealed that this model accounted for approximately 43% of the variance between risk levels and was an effective discriminator for risk-level classification ($\chi^2 = 25.401$, $p = .000$). Using this model, it was predicted that 75% of normal and 88% of at-risk students were classified correctly. *Model 2* was statistically driven, using the stepwise method to determine which of the three composites were the most salient discriminators for risk-level classification. The resulting model included only the Academic Achievement and Engagement composites and accounted for approximately 43% of the variance between risk levels. Similar to the previous model, it was also deemed an effective discriminator for risk-level classification ($\chi^2 = 25.495$, $p = .000$). Furthermore, using this latter model, it was predicted that 67% of normal and 88% of at-risk students were classified correctly. See Table 3 and Table 4 for summaries of these results.

TABLE 1: *Correlations between BESS Risk Level and Report Card Outcomes*

Report Card Outcomes	Correlation with Risk Level	<i>p</i> -value
<i>Academic Sub-Composites</i>		
Listening	-.503	.000
Reading	-.461	.000
Writing	-.435	.001
Math	-.393	.003
History	-.528	.000
Science	-.674	.000
<i>General Composites</i>		
Academic Achievement	-.549	.000
Engagement (Behavioral)	-.614	.000
Behavioral Performance	-.507	.000

TABLE 2: *One-Way ANOVA for BESS Risk Level and Report Card Outcomes*

Report Card Outcomes	<i>F</i>	<i>p</i> -value
Between Groups		
<i>Academic Sub-Composites</i>		
Listening	15.545	.000
Reading	12.440	.001
Writing	10.745	.002
Math	8.424	.006
History	17.769	.000
Science	20.000	.000
<i>General Composites</i>		
Academic Achievement	19.827	.000
Engagement (Behavioral)	27.824	.000
Behavioral Performance	15.947	.000

TABLE 3: *Discriminant Function Analyses: Model Validity*

Variables	Wilk's λ	χ^2	<i>p</i> -value
<i>Model 1</i> (Theoretical)* Academic Achievement Engagement Behavioral Performance	.565	25.401	.000
<i>Model 2</i> (Statistical)** Step 1: Engagement Step 2: Academic Achievement	.567	25.495	.000

*All variables were entered simultaneously.

** Variables entered using stepwise method.

TABLE 4: *Discriminant Function Analyses: Group Membership Classifications*

Group	Predicted Group Membership	
	Normal	At-Risk
<i>Model 1</i> (Theoretical)*		
Normal	18 (75%)	6 (25%)
At-Risk	3 (12%)	21 (88%)
<i>Model 2</i> (Statistical)**		
Normal	16 (67%)	8 (33%)
At-Risk	3 (12%)	21 (88%)

* 80.1% of original grouped cases correctly classified.

** 77.1% of original grouped cases correctly classified.

DISCUSSION

The purpose of the present study was to further examine the concurrent validity of the BESS in elementary school settings, by examining the relation between teachers' BESS ratings and recent report-card outcomes. Consistent with the extant research, the results supported our hypotheses that students' risk-level classifications would be significantly related to school-based outcome criteria and that such school-based outcome criteria, as reported via report cards, would be effective discriminators of students' risk-level classification.

These results provide additional concurrent validity evidence for using the BESS in elementary school settings, showing that teacher ratings on the screener are highly related to their evaluations of students' academic, engagement, and behavioral outcomes, as reported via report cards. Specifically, results revealed that BESS risk-level classification was moderately or highly correlated with all of the academic sub-composites as well as the three general composites; that there were significant mean differences between normal and at-risk students on all sub-composites and composites; and that the report-card derived composites were effective discriminators for risk-level classification. Interestingly, however, results also suggest that similar discriminant results could be obtained by excluding the Behavioral Performance composite. Thus, such findings warrant further examination of which school-based criteria will provide the best concurrent validity for the BESS. But in general, these findings add to the existing evidence supporting the BESS as an appropriate school-based screening instrument.

Limitations

Although the results of the present study supported the concurrent validity of the BESS, there are three limitations. First, the sample characteristics were limited: only third- and fourth-graders participated in this study; the students were all from the same district; and they identified as predominantly Hispanic or White. Thus, the results have low generalizability for students in other grades or locations and identifying with other ethnicities. Second, this study, similar to the previous research validating the BESS, focused only on teacher ratings. Given that the BESS also consists of parent- and self-rating components, it is unknown how these school-based criteria would relate to risk-level classifications derived from other raters. As such, these results should only be construed as supporting one facet of the overall BESS. Third, the school-based criteria used herein—report card outcomes—are idiosyncratic to the local school district. Other districts within other states have varying report-card indicators; thus, further replication is warranted.

Future Directions

Despite limitations, findings from this study suggest several directions for future research. Foremost, the significant relations observed between BESS risk-level classification and local report-card criteria warrant further examination with various other report-card indicators, across grade levels, within varying school contexts, and using other behavioral indicators. To enhance the criterion-related validity, such outcomes could be used for both concurrent and predictive evidence. Furthermore, the difference in validity coefficients between local school criteria (e.g., report card grades) and global school criteria (e.g., standardized testing results and other BASC-2 ratings) also warrants evaluation. Previous validation studies (DiStefano & Kamphaus, 2007; Kamphaus et al., 2007) have focused primarily on global school criteria, providing limited information on the criteria inherent within the local school context. And lastly, future research may also benefit from utilizing further discriminant analyses, attempting to determine which types of criteria are the most effective for discriminating between students' risk-level classification.

Implications for Practice

Considering the results of the present study in conjunction with the existing analyses of the BESS, the results are promising. The psychometric qualities and school-based validity of the BESS appear suitable for using the instrument as a universal screener for students, seeking to identify those at-risk for potential emotional or behavioral problems. As noted in the introduction, it is essential that universal screening should always be (a) integrated within a larger student-support framework, (b) accompanied by well-defined objectives, and (c) be established through careful consideration and planning (Dowdy et al., in press). By incorporating these elements and implementing a valid screening instrument like the BESS, school psychologists may ultimately enhance early identification and intervention efforts within their local school context – preventing the development of more severe problems and promoting more positive outcomes for students in the future.

REFERENCES

- Caldarella, P., Young, E.L., Richardson, M.J., Young, B.J., & Young, K.R. (2008). Validation of the Systematic Screening for Behavior Disorders in middle and junior high school. *Journal of Emotional and Behavioral Disorders, 16*, 105-117.
- DiStefano, C.A., & Kamphaus, R.W. (2007). Development and validation of a behavioral screener for preschool-age children. *Journal of Emotional and Behavioral Disorders, 15*, 93-102.
- Dowdy, E.T., Furlong, M.J., Eklund, K., Saeki, E., & Ritchey, K. (in press). Screening for mental health and wellness: Current school-based practices and emerging possibilities. In B. Doll (Ed.), *Handbook of prevention science*. Bethesda, MA: National Association of School Psychologists.
- Glover, T., & Albers, C. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45*, 117-135.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology, Psychiatry, and Allied Disciplines, 38*, 581-586.
- Kamphaus, R.W., & Reynolds, C.R. (2007). *BASC-2 Behavioral and Emotional Screening System Manual*. Circle Pines, MN: Pearson.
- Kamphaus, R.W., Thorpe, J.S., Winsor, A.P., Kroncke, A.P., Dowdy, E.T., & VanDeventer, M.C. (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. *Educational and Psychological Measurement, 67*, 1-15.
- Little, M., Murphy, J.M., Jellinek, M.S., Bishop, S.J., & Arnett, H.L. (1994). Screening 4- and 5-year-old children for psychosocial dysfunction: A preliminary study with the Pediatric Symptom Checklist. *Journal of Developmental and Behavioral Pediatrics, 15*, 191-197.
- Reynolds, C.R., & Kamphaus, R.W. (2004). *Behavior Assessment System for Children-2*. Circle Pines, MN: Pearson
- Romer, D., & McIntosh, M. (2005). The roles and perspectives of school mental health professionals in promoting adolescent mental health. In D.L. Evans, E.B. Foa, R.E. Gur, H. Hendin, C.P. O'Brien, M.E.P. Seligman, & B.T. Walsh (Eds.), *Treating and preventing adolescent mental health disorders: What we know and what we don't know* (pp. 598-615). New York: Oxford University Press.
- Rones, M., Hoagwood, K. (2000). School-based mental health services: A research review. *Clinical Child and Family Psychology Review, 3*, 223-241.
- Walker, H.M., & Severson, H.H. (1992). *Systematic Screening for Behavior Disorders* (2nd ed.). Longmont, CO: Sopris West.