

Screening for Early Learning Problems within an urban Population: The Brief Academic Competence Evaluation Screening System

Ryan J. Kettler

Peabody College at Vanderbilt University

The Brief Academic Competence Evaluation Screening System (BACESS; Elliott, DiPerna, & Huai, 2003) is a multi-phase instrument designed to assist educators in the identification of students who are likely to experience early learning problems. The BACESS was used in eight elementary classrooms (n = 71) in southern California. Each phase of the BACESS was found to be highly reliable, and the BACESS was found to share concurrent validity with the California Standards Tests. Teacher feedback via an evaluation survey indicated that phases 1 and 2 of the system were time efficient and useful.

Educators have been screening elementary school students for future academic and behavior problems since the 1940's (Gredler, 1997). The rationale behind screening for future academic problems is based on the theory that special learning needs, analogous to medical diseases, progress linearly and become worse over time (Severson & Walker, 2002). If special learning needs can be identified earlier, educators have a better chance of intervening and correcting problems before they become pervasive. If properly developed and validated, screening systems that are linked to quality interventions can reduce referrals to special education and facilitate an identification process that is proactive.

Demand for screening instruments has increased over the past 50 years, both because of a growth in the number of intervention programs available for at-risk students, and because of legislation that includes greater accountability for academic failure. In 2001, Congress passed the No Child Left Behind Act (NCLB), which indicated that universal screening systems for reading should be adopted in order to help low-achieving students meet high academic standards.

In addition, a recent report from the National Research Council (NRC, Donovan & Cross, 2002) recommended that states utilize universal screening methods for reading and behavior problems in order to improve the early identification of students at-risk for academic difficulties. The report indicated that universal screening could help correct problems such as disproportionate minority representation in special education and the gap between academic assessment and intervention. The NRC recommended that screening systems should: incorporate multiple tiers, be developed with input from large-scale research centers, and be implemented at a federal level. An accurate and practical screening system for early identification of special learning needs would meet these criteria.

Teacher Ratings of Academic Performance

Teacher ratings are one relatively accurate and cost effective method of evaluating students' learning abilities. Gerber and Semmel (1984) came to this conclusion after reviewing a decade of literature. They noted that teachers generate the initial referral for most potentially at-risk students, and that approximately 70% of students whom teachers refer are eventually classified with a learning disability. The authors attributed this high success rate to the fact that teachers have daily contacts with students, and have a meaningful context in which to evaluate students' performance. Other researchers have obtained findings consistent with those of Gerber and Semmel (1984). Two

studies published by Gresham and colleagues (Gresham, MacMillan, & Bocean, 1997; Gresham, Reschly, & Carey, 1987) indicated a high concurrence between special education recommendations and teacher opinions of academic ability.

In the earlier study (Gresham et al., 1987), teachers confirmed that 96% of students diagnosed with a learning disability indeed had a learning disability. In the later study (Gresham et al., 1997), teachers agreed with the diagnoses of 91% of students with learning disabilities, 95% of students exhibiting low achievement, and 100% of students with low IQ's. Demaray and Elliott (1998) asked teachers to rate student performance via the Academic Competence Scale of the Social Skills Rating System – Teacher Form (SSRS-T; Gresham & Elliott, 1990). The correlation between teachers' evaluations via the SSRS-T and students' academic achievement scores on the Kaufman Test of Educational Achievement, Brief Form (K-TEA; Kaufman & Kaufman, 1985) was moderately high ($r = .70$). Flynn and Rahbar (1998) found that teachers who were provided a 29-item rating scale were able to evaluate students' academic performance much more accurately than when asked to informally identify students that were struggling academically. Collectively this research indicates that teacher ratings are an acceptable method for identifying students who may have early learning problems.

Brief Academic Competence Evaluation Screening System (BACESS)

The BACESS is a screening instrument based on teacher ratings that can fill the role described in the report from the NRC (2002) and NCLB (2001), by helping to identify students who are at-risk for academic failure at an early age. The BACESS (Elliott, DiPerna, & Huai, 2003) was conceptualized as an outgrowth of the Academic Competence Evaluation Scales (ACES; DiPerna & Elliott,

2000), a set of rating scales that measure student academic skills and enablers. As presently conceptualized, the BACCESS is a three-phase system involving teacher nominations of struggling students, teacher ratings based on grade level expectations, and comprehensive ratings against national norms. During Phase 1, teachers nominate all students in their class based on comprehensive scoring rubrics for reading, language arts, math, and social behavior, into one of five different levels.

In Phase 2, teachers rate students passing through Phase 1 on five key academic skills and five key academic enablers. Academic skills are the content specific skills (e.g., uses numbers to solve daily problems) that help students perform in particular subjects, while academic enablers (e.g., participates in class discussions) are attributes that help students in all academic areas. In Phase 3, teachers complete the entire ACES for students who advance through Phase 2, in order to obtain nationally normed scores for academic skills and academic enablers.

The current study is part of a line of research developed to evaluate the reliability and validity of the BACCESS in multiple educational settings, in order to determine whether the instrument is appropriate for use on a state- or district-wide basis. In one study involving 25 teachers and 285 students in Wisconsin, Phases 1 and 2 of the BACCESS were found to have high reliability coefficients within the context of relatively short screening tools (Kettler, Elliott, & Albers, 2007).

The two phases together were found to have sensitivity (.67) and specificity (.80) comparable to other academic screening instruments, when achievement proficiency tests were used as an outcome measure. While this study provided promising evidence for the BACCESS in a primarily European American (88%) and high achieving population (78% attained proficiency in reading, language arts, and mathematics), it remains to be determined whether the instrument would perform as well in other settings.

Conceptual Framework

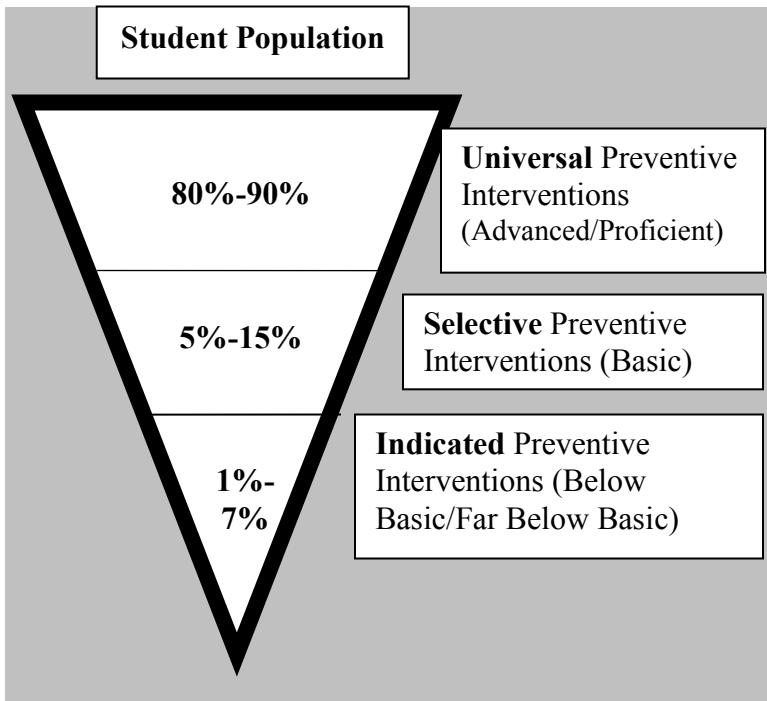
The conceptual framework for evaluating the BACESS and many academic screening instruments is based on three main theoretical foundations. The first foundation is that students experience a continuum of preventive intervention needs, as described by the three-tiered framework (Larson, 1994). The second foundation is that educational and psychological assessment tools should be evaluated via a systematic process, informed by the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), that addresses the reliability, validity, and utility of an instrument for its intended purpose. The third theoretical foundation is that because the purpose of a screening instrument is to identify the early stages of a problem, rather than to measure a construct, Bayesian conditional probability analyses are a helpful way to characterize concurrent validity (Bennett et al., 1999).

Walker and Shinn (2002) wrote that students who are identified as struggling should be provided interventions conceptualized within Larson's (1994) three-tiered model commonly used within the public health domain. The tiers of the model correspond to three categories of students identified by Walker and Shinn (2002): (a) typically developing students, (b) students who are at elevated risk, and (c) students who show signs of life-course persistent difficulties.

Gordon (1987) suggested using the terms *universal*, *selective*, and *indicated* to describe this spectrum of prevention and intervention needs, and also provided the term preventive intervention. In 2001, the National Institute of Mental Health (NIMH) adopted these terms, rather than primary, secondary, and tertiary, due to the perception that they more accurately captured the preventive nature of interventions at all three levels.

In the state of California, the terms Advanced, Proficient, Basic, Below Basic, and Far Below Basic are used to describe student performances in each academic content area during proficiency testing. Figure 1 depicts the hypothesized relationships between the model of universal, selective, and indicated preventive interventions (Gordon, 1987; NIMH, 2001), and the model used in the current study based on proficiency test results in California.

FIGURE 1. *Hypothesized Relationship between the Three-Tier Model and Proficiency Test Results*



The key concern with any assessment tool is whether its scores are valid representations of the constructs that they are intended to measure. However, before an instrument can be proven valid for a purpose, it must be proven reliable. A reliable tool is one that measures or identifies the same trait

consistently.

Reliability is estimated by calculating the equivalence of scores attained from a measure under conditions that should produce relatively equivalent scores (e.g., the same rater and a stable construct at different times, two subsets of items from the same scale at the same time, two raters of one construct at the same time, etc.). One way to estimate reliability for measures that have a small number of items is to calculate the correlation between each item and the total of all of the other items on the scale (Walker et al., 1988). This estimate of reliability is referred to as item-total reliability, and has the advantage of being exempt from error due to data being collected at different points in time, or being submitted by different raters.

A similar method of characterizing reliability, coefficient alpha, indicates how well a larger set of items fit together to measure a single construct. All of the aforementioned advantages of item-total correlations apply to coefficient alpha. Once evidence of an instrument's reliability has been established, the issue of whether it has construct validity for its intended purpose can be considered.

Construct validity is the degree to which an instrument measures that which it is intended to measure, or identifies that which it is intended to identify. Validity based on relationships with other variables is one form of construct validity evidence mentioned in the *Standards for Educational and Psychological Testing* (American Educational Research Association, 1999) that is very important for evaluating a screening system, because a screening system is intended to discriminate between the presence and absence of a condition (e.g., the presence or absence of early learning problems). Validity based on relationships with other variables can be further classified by whether the two variables are measured at the same time (i.e., concurrent validity) or at different times (i.e., predictive validity). The focus of this study is concurrent validity.

One method of characterizing the accuracy of screening

systems for dichotomous outcomes has gained popularity in educational sciences: Bayesian conditional probability or sensitivity/specificity analyses (Bennett et al., 1999). Bayesian conditional probability analyses require an independent variable that screens students into two different classifications (i.e., early learning problems vs. the absence of early learning problems) and a dependent variable that serves as a “gold standard” and also divides students into two classifications (i.e., students who are experiencing early learning problems and students who are not). The analyses are based on the fact that the combination of the two variables yields four possible outcomes: (a) a student may be screened and identified as having an early learning problem and be actually experiencing the early stages of a learning problem, (b) a student may be screened and identified as having an early learning problem but not actually be experiencing the early stages of a learning problem, (c) a student may be screened and not identified as having an early learning problem but actually be experiencing the early stages of a learning problem, or (d) a student may be screened and not identified as having an early learning problem and not be experiencing the early stages of a learning problem (Bennett et al., 1999).

The possible outcomes are depicted in Table 1. From these four outcomes, a screening system can be evaluated on the following Bayesian conditional probability indices: sensitivity (the likelihood that a screener will correctly identify a need), specificity (the likelihood that a screener will correctly not identify a need), positive predictive power (the likelihood that an identified student is one that has a need), and negative predictive power (the likelihood that a student who is not identified is one who does not have a need). This framework can be useful for evaluating the decision rules or cutoff scores of a measure because it accurately reflects how an increase on any one of these indices tends to co-occur with a decrease on another.

TABLE 1 *Possible Outcomes within a Bayesian Conditional Probability Framework*

		Eventual Outcome	
		Early Learning Problem	No Early Learning Problem
Screening Indicator	At-Risk	a	b
	Not At-Risk	c	d

Note: Sensitivity = $a/(a+c)$; specificity = $d/(b+d)$; positive predictive power = $a/(a+b)$; and negative predictive power = $d/(c+d)$. This Figure is adapted from Bennett, et al. (1999).

Research Questions

The BACESS has been shown to be a reliable and accurate screening instrument when used with a primarily European American, Midwestern sample with a relatively low base rate of academic learning problems (Kettler et al., 2007). The current study was designed to replicate previous findings in an urban sample of primarily Latino American students with a relatively high base rate of academic learning problems. The following research questions were inspired by the need for a reliable, valid, and useful broadband academic screening system:

- *Is the BACESS a reliable predictor of early learning problems?*
- *Is the BACESS a valid predictor of early learning problems?*
- *Do teachers find the BACESS and each of its phases useful?*

Method

Participants

Participants in the current study included teachers and students from eight classrooms in an urban elementary school in southern California. All eight teachers in the sample were female, including four European Americans, two Asian Americans, and one African American (one teacher did not report ethnicity). Teachers in the study taught classrooms that included a mean of 19.14 (*S.D.* = 0.99) students. The student sample was 60% female and 94% Latino American. The sample included 39 first grade students, 18 second grade students, and 14 third grade students. Concurrent validity analyses were performed on the subset of 27 second and third grade students who participated in proficiency testing during the year that the BACESS was administered.

Data Collection Procedures

Brief Academic Competence Evaluation Screening System. Under evaluation, the BACESS is described in detail in the Introduction section of this manuscript.

Evaluation Survey. The Evaluation Survey was based on an instrument developed by Huai (2004) for users of the BACESS to provide feedback regarding the instrument. It includes seven questions related to the instrument answered on a four-point Likert scale (1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Agree*, and 4 = *Strongly Agree*), prompts for teachers to estimate how much time they spent on the first two phases of the BACESS, and provides opportunities for brief written responses to open questions.

Background Information Questionnaire. The Background Information Questionnaire was completed by teachers to share details on their demographics and experiences. It includes questions about gender and ethnicity, as well as grade level, classroom size, and previous

experiences with teaching, pre-referral intervention, and the ACES system.

California Standards Tests. The California Standards Tests (CST's) were designed to measure students' mastery of content standards in English-Language Arts and Mathematics. Both tests consist of 65 multiple choice questions. Previous versions of the test have been found to be highly reliable and valid for measuring students' mastery of California content standards (California Department of Education, 2003, 2005).

Procedure

Teachers participated by completing the BACESS, the Evaluation Survey, and the Background Information Questionnaire. Students participated by completing the CST's in Mathematics and Reading administered to all students in grades 2 through 11 in California. Although completion of the BACESS and the CST's occurred during the same time period, it is important to note that teachers did not have access to the results of the CST's when completing the screening system.

Data Analysis

The reliability of Phase 1 of the BACESS was characterized by calculating the correlations between scores in each content area (Reading, Language Arts, Mathematics, and Social Behavior) and a sum score based on all four areas. The reliability of Phase 2 was characterized by calculating Cronbach's alpha. Concurrent validity between the BACESS and the CST's was characterized via Bayesian conditional probability indices taken by second and third grade students. The utility of the BACESS was evaluated via descriptive quantitative and qualitative analysis of Evaluation Survey answers. The reliability and validity of Phase 3 were already well established and documented in the ACES manual (DiPerna & Elliott, 2000).

Results and Discussion

Reliability coefficients for phase 1, estimated by the correlations of each individual subscale with a total score from the sum of subscales, ranged between $r = .81$ and $r = .91$. Cronbach's alpha for Phase 2 was $r = .94$. These reliabilities are quite high considering they refer to short phases from a multi-stage screening instrument.

Concurrent validity evidence indicated adequate performance for the BACESS with regard to agreement with the CST's. Table 2 depicts agreement between the BACESS and proficiency as determined via the CST's within the subsample of students in 2nd and 3rd grade. The BACESS was sufficiently specific (.68) and sensitive (.60) for this purpose, and the positive predictive power (.88) of the instrument was quite high. This finding indicates that the BACESS identified almost exclusively students who needed help, reducing the possibility that resources would be wasted. The quality of the instrument as characterized by negative predictive power (.30) was low, indicating that too high of a proportion of students who did not qualify via that BACESS actually were showing signs of early learning problems. The danger associated with such low negative

Table 2 *California Standards Tests Proficiency Level by BACESS Qualification*

		Eventual Outcome	
		Early Learning Problem	No Early Learning Problem
Screening Indicator	At-Risk	15	2
	Not At-Risk	7	3

Note: Sensitivity = .68; specificity = .60; positive predictive power = .88; and negative predictive power = .30.

predictive power is that too many students with learning problems remain unidentified.

Via the evaluation survey, teachers indicated having spent an average of 54 minutes (*S.D.* = 27 minutes) completing the first two phases of BACCESS. On average, 23 minutes (*S.D.* = 13 minutes) was spent using Phase 1 and 31 minutes (*S.D.* = 16 minutes) was spent using Phase 2. The amount of time spent completing Phase 3, obtaining nationally normed academic skills and enablers scores that link to prereferral interventions, was previously determined to be less than 20 minutes per individual (DiPerna & Elliott, 2000). All eight teachers agreed that the time spent on Phases 1 & 2 was reasonable, and that the BACCESS as a whole was useful. Only four of seven teachers indicated that the BACCESS is easy to use, primarily citing Phase 3 as being the most difficult.

Implications and Conclusion

The BACCESS is a highly reliable screening instrument when used within an urban, primarily Latino American elementary school environment. As currently constructed, the instrument appears to be related to achievement test proficiency scores (i.e., the CST's), although negative predictive power is low. More evidence needs to be collected, with a larger sample and of a predictive nature, before results of the current study can be generalized with confidence.

While teachers see Phases 1 and 2 as time efficient, helpful, and useful, Phase 3 is seen as difficult to complete, and may be more appropriate for professionals with a background in psychological assessment to interpret. The practical implications of the current findings are that teacher ratings organized within a screening system appear to be an acceptable method for identifying early learning problems. Previously shown to be an accurate screening system for academic problems within a relatively high achieving

population (Kettler et al., 2007), the BACESS may also be an acceptable screening tool for use in an environment in which the majority of students (81% in the current sample) do not obtain proficiency across reading and mathematics.

This finding is important because variations in base rate affect the attainability of high scores on the various Bayesian conditional probability indices. In the current sample, a high proportion of the students identified by the screening system are ideal candidates for selective preventive intervention prior to referral for special education. Future revisions of the BACESS should be aimed at maintaining this level of success, while decreasing the rate of false negative cases among academically challenged populations.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D.C.: Author.
- Bennett, K.J., Lipman, E.L., Brown, S., Racine, Y., Boyle, M.H., & Offord, D.R. (1999). Predicting conduct problems: Can high-risk children be identified in kindergarten and grade 1? *Journal of Consulting and Clinical Psychology*, *67*, 470-480.
- California Department of Education. (2003). *Accountability workbook attachment S – California's consolidated state application*. Retrieved April 1, 2007 from <http://www.cde.ca.gov/nclb/sr/sa/documents/yr03wb0131atts.pdf>.
- California Department of Education. (2005). *December 2005 Informational Memorandum*. Retrieved April 1, 2007 <http://www.cde.ca.gov/be/pn/im/infoabsaddec05itm03.asp>.

- Demaray, M.K., & Elliott, S.N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performance. *School Psychology Quarterly, 13*, 8-24.
- DiPerna, J.C., & Elliott, S.N. (2000). *Academic Competence Evaluation Scales manual K-12*. San Antonio, TX: The Psychological Corporation.
- Donovan, M.S., & Cross, C.T. (2002). *Minority students in special and gifted education*. Washington, D.C.: National Academy Press.
- Elliott, S.N., DiPerna, J.C., & Huai, N. (2003). *Brief Academic Competence Evaluation Screening System*. Unpublished instrument. University of Wisconsin-Madison.
- Flynn, J.M., & Rahbar, M.H. (1998). Improving teacher prediction of children at risk for reading failure. *Psychology in the Schools, 35*, 163-172.
- Gerber, M.M., & Semmel, M.I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. *Educational Psychologist, 19*, 137-148.
- Gordon, R. (1987). An operational classification of disease prevention. In J.A. Steinberg, & M.M. Silverman (Eds.), *Preventing mental disorders: a research perspective*. Rockville, MD, Department of Health and Human Services: 20-26.
- Gredler, G.R. (1997). Issues in early childhood screening and assessment. *Psychology in the Schools, 34*, 99-106.
- Gresham, F.M., & Elliott, S.N. (1990). *The Social Skills Rating System*. Circle Pines, MN: American Guidance Service.
- Gresham, F.M., MacMillan, D.L., & Bocian, K.M. (1997). Teachers as "tests": Differential validity of teacher judgments in identifying students at-risk for learning disabilities. *School Psychology Review, 26*, 47-60.

- Gresham, F.M., Reschly, D.J., & Carey, M.P. (1987). Teachers as "tests": Classification accuracy and concurrent validation on the identification of learning disabled children. *School Psychology Review*, 16, 543-553.
- Huai, N. (2004). *Development and validation of a universal screening instrument: The Brief Academic Competence Evaluation Screening System for identifying children at-risk of school failures*. Unpublished dissertation: University of Wisconsin-Madison.
- Kaufman, A.S., & Kaufman, N.L. (1985). Kaufman Test of Educational Achievement Brief Form. Circle Pines, MN: American Guidance Service.
- Kettler, R.J., Elliott, S.N., & Albers, C.A. (in press). Structured teacher ratings to identify students who need help: Validation of the Brief Academic Competence Evaluation Screening System. *Journal of Psychoeducational Assessment*.
- Larson, J. (1994). Violence prevention in the schools: A review of selected programs and procedures. *School Psychology Review*, 23, 151-164.
- National Institute of Mental Health. (2001). National Advisory Mental Health Council Workgroup on Mental Disorders Prevention Research. Priorities for prevention research at NIMH. *Prevention & Treatment*, 4, NP. (Posted June 26, 2001).
- Severson, H., & Walker, H. (2002). Pro-active approaches for identifying children at risk for socio-behavioral problems. In K.L. Lane, F.M. Gresham and T. O'Shaughnessy (Eds.), *Interventions for students with or at-risk for emotional and behavioral disorders*. (pp. 33-54). Boston, MA: Allyn and Bacon.

- Walker, H.M., Severson, H., Stiller, B., Williams, G., Haring, N., Shinn, M., & Todis, B. (1988). Systematic screening of pupils in the elementary age range at risk for behavioral disorders: Development and trial testing of a multiple gating model. *Remedial and Special Education, 9* (3), 8-14.
- Walker, H.M., & Shinn, M. (2002). Structured school-based interventions to achieve integrated primary, secondary, and tertiary prevention goals for safe and effective schools. In M.R. Shinn, H.M. Walker, & G.Stoner, Eds.), *Interventions for academic and behavior problems II: Preventative and remedial approaches* (pp. 1-25). Bethesda, MD: National Association of School Psychologists.
- U.S. Department of Education. (2001). *No child left behind*. Retrieved August 21, 2001, from <http://www.ed.gov/inits/nclb/titlepage.html>.