Exceptional Children

Vol. 75, No. 3, pp. 321-340.

An Examination of the Evidence Base for Function-Based Interventions for Students With Emotional and/or Behavioral Disorders Attending Middle and High Schools

KATHLEEN LYNNE LANE
JEMMA ROBERTSON KALBERG
JENNA COURTNEY SHEPCARO
Peabody College of Vanderbilt University

ABSTRACT: *The authors field-tested the core quality indicators and standards for evidence-based practices for single-case design studies developed by Horner and colleagues (2005) by applying them to the literature exploring functional assessment-based interventions conducted with secondary-age students with emotional and/or behavioral disorders (EBD). First, we evaluated this knowledge base by applying the indicators to determine if the studies identified (*n = *12) were of acceptable methodological quality. Second, we analyzed studies meeting the recommended quality indicators to determine whether function-based interventions with students with EBD might be considered an evidence-based practice. Results reveal that only 1 study addressed all proposed quality indicators, suggesting that function-based interventions are not yet an evidence-based practice for this population per these indicators and standards. Limitations and recommendations are posed.*

Students with emotional and/or behavioral disorders (EBD) represent between 2% and 20% of the school-age population and are among some of the most challenging students to teach (Walker, Ramsey, & Gresham, 2004). By definition, students with EBD have behavioral, social, and academic deficits that pose challenges within and beyond the school setting (Kauffman, 2005). For example, they have impaired social skills that strain relationships with teachers and peers (Gresham, 2002). In addition, students with EBD have broad academic deficits that, at best, remain

stable over time (Nelson, Benner, Lane, & Smith, 2004). Unfortunately, outcomes do not improve when EBD students leave the school setting as evidenced by employment difficulties, contact with the juvenile justice system, limited community involvement, and high rates of access to mental health services (Bullis & Yovanoff, 2006).

During the past 30 years, schools have responded with a range of interventions to support these youngsters including schoolwide primary prevention efforts (e.g., antibullying programs); secondary prevention efforts (e.g., small group instruction in conflict resolution skills); and tertiary prevention efforts (e.g., individualized intervention efforts; Horner & Sugai, 2000). One tertiary intervention effort that has met with demonstrated success, particularly with elementary-age students with EBD is function-based interventions (Conroy, Dunlap, Clarke, & Alter, 2005; Kern, Hilt, & Gresham, 2004; Lane, Umbreit, & Beebe-Frankenberger, 1999).

Function-based interventions refer to interventions designed based on the reasons why problem behaviors occur (Umbreit, Ferro, Liaupsin, & Lane, 2007). The motive for a given behavior is derived through a functional behavioral assessment. In brief, descriptive (e.g., interviews, direct observations of behavior, rating scales) and experimental (e.g., functional analysis) procedures are used to identify the antecedent conditions that prompt a target behavior (e.g., disruption) to occur and the consequences that maintain the behavior. These data are used to generate a hypothesis statement regarding the function of the behavior.

In general, all behaviors occur to either obtain (positive reinforcement) or avoid (negative reinforcement) attention; activities or tasks; or tangible or sensory conditions (Umbreit et al., 2007). Often, the hypothesis statement is tested by systematically manipulating environmental conditions to identify or confirm maintaining consequences. Next, an intervention is designed based on the function of the target behavior with a goal of teaching the student a more reliable, efficient method of meeting his or her objective (e.g., escaping a too difficult or too easy task; Umbreit, Lane, & Dejud, 2004). This is done by constructing an intervention that (a) adjusts antecedent conditions that prompt the problem behavior, (b) increases reinforcement rates for the replacement behavior, and (c) extinguishes reinforcement for the target behavior.

Functional assessment procedures were originally developed in clinical settings with individuals with developmental disabilities (Iwata, Dorsey, Slifer, Bauman, & Richman, 1982). Since that time, functional assessment-based interventions have been used to shape a variety of behaviors in a range of educational settings (e.g., general education classes, self-contained classrooms, self-contained schools), with students with a range of conditions including severe disabilities (Sasso, Reimers, Cooper, & Wacker, 1992); attention deficit disorders and behavioral concerns (Ervin, DuPaul, Kern, & Friman, 1998); and emotional and/or behavioral problems (Kern, Childs, Dunlap, Clarke, & Falk, 1994; Kern, Delaney, Clarke, Dunlap, & Childs, 2001).

In fact, functional behavioral assessments have been endorsed by the National Association of School Psychologists, National Association of State Directors of Education, and National Institutes of Health, and mandated in the Individuals With Disabilities Education Act (IDEA; first in 1997 and again in 2004) when certain disciplinary circumstances occur (Kern et al., 2004). Namely, school personnel must conduct a functional behavioral assessment when (a) a student is placed in an alternative placement for behavior deemed to be dangerous to self or others; (b) a student is placed in an alternative setting for 45 days due to drug or weapons violations; or (c) a student's suspension or alternative setting placement extends beyond 10 days or constitutes a change in placement (Drasgow & Yell, 2001). Given the behaviors typical of students with EBD, many of these students may require function-based interventions.

Yet, several researchers contend that such a mandate may not be entirely appropriate (e.g., Fox, Conroy, & Heckaman, 1998; Gresham, 2004; Kern et al., 2004; Quinn et al., 2001; Sasso, Conroy, Stichter, & Fox, 2001). Specifically, there are concerns of a generalization error in the sense that existing functional assessment procedures, which were originally developed for persons with developmental disabilities, have not been validated for use with students with EBD (Fox et al.; Kern et al., 2004; Sasso et al., 2001). At best, there is a

modest body of literature exploring the effectiveness of function-based interventions for students with EBD, with most of the studies conducted in elementary grades (Lane et al., 1999; Quinn et al.). In the reviews of function-based interventions conducted with students with and at risk for EBD, the populations have been predominantly male, with limited inquiry with secondary-age students (Conroy et al., 2005; Kern et al., 2004; Lane et al., 1999; Sasso et al., 2001).

Therefore, questions arise as to the efficacy of function-based interventions, particularly for older school-age students. It is possible that designing, implementing, and evaluating function-based interventions with this population will prove to be a highly formidable task given the increased importance of the peer group (Morrison, Robertson, Laurie, & Kelly, 2002); topographical changes in discipline problems (e.g., covert acts of aggression, internalizing behaviors; Loeber, Green, Lahey, Frick, & McBurnett, 2000; Morris, Shah, & Morris, 2002); and difficulties in identifying meaningful reinforcers that can compete with the reinforcing value of the undesired, target behavior (e.g., truancy). Thus, the question arises: Are functional assessment-based interventions an evidence-based practice for secondary-age students with EBD?

Answers to questions about intervention efficacy have become more complex as researchers have sought to define what constitutes evidence-based practices. Gersten et al. (2005) and Horner et al. (2005) introduced criteria for determining whether a practice is evidence-based using group design and single-case experimental investigations, respectively. These research teams developed quality indicators for group design and single-case design inquiry that can be used to determine the extent to which a given study meets requisite criteria, thereby establishing the study as a reputable, appropriate study. Further, each team offered guidelines for evaluating bodies of reputable studies that meet the quality indicators to determine if the practice is evidence based.

The goal of this review was to field-test these quality indicators by applying them to the body of literature exploring functional assessment-based interventions conducted with secondary-age students with EBD. Specifically, the intent was threefold. First, given that this body of literature focused on single-case methodology, we evaluated this knowledge base by field-testing the quality indicators posed by Horner et al. (2005) to determine if the studies identified in a systematic literature review met the recommended quality indicators. Second, we analyzed studies that met the recommended quality indicators to determine whether function-based interventions with secondary-age students with EBD are an evidence-based practice according to Horner et al.'s proposed standards. Third, we discussed the extent to which the quality indicators represent reasonable standards and offered considerations for future application and evaluation of the quality indicators.

Are functional assessment-based interventions an evidence-based practice for secondary-age students with EBD?

METHOD

Article Selection Procedures

We conducted a systematic search of psychology and educational databases (PsycINFO and Educational Resources Information Center, ERIC) to identify function-based intervention studies conducted with secondary-age students with or at risk for EBD. Search terms included all possible combinations and derivatives of the following sets of terms: (a) *functional assessment, functional analysis, assessment based, intervention,* and *procedures;* and (b) *seriously emotionally disturbed, emotional and/or behavioral disorders, at risk,* and *problem behavior* (Lane et al., 1999). The title and abstract of each article from the electronic search was evaluated to determine if the article should be read in its entirety to evaluate inclusion eligibility. Next, a master list of journals that published the included studies was created. We conducted hand searches of those journals that published two or more of the articles from 1980 to present to gather any other articles that met inclusion criteria. Searches were conducted in the following journals: *Behavioral Disorders, Education and Treatment of Children, Journal of Applied Behavior Analysis, Journal of Emotional and Behavioral Disorders, Journal of*

Positive Behavior Interventions, and *School Psychology Review*. Finally, we compared our search results with other reviews of function-based interventions (e.g., Dunlap & Childs, 1996; Heckaman, Conroy, Fox, & Chait, 2000; Lane et al., 1999).

Thirty-three articles, all of which employed single-subject designs, were identified as appropriate for further review using the procedures previously stated. Each article was read in its entirety to determine if the article met the following inclusion criteria.

Inclusion Criteria

The intent of this review was to evaluate the extent to which function-based interventions conducted with secondary-age students with or at risk for EBD met the recommended indicators and to determine if function-based interventions are an evidence-based practice for this population. Studies were included in this review only if (a) the participants were diagnosed with or were at risk for EBD, (b) the participants were educated in a secondary school setting, (c) an intervention derived from a functional assessment was implemented and evaluated using single-case methodology, (d) intervention results included a graphic display of student outcomes, and (e) the study was published in a refereed journal.

Participants included in the studies had to be adolescents, defined as students ages 13 to 18, with or at risk for EBD. This group included students with

- EBD, an inclusive term to describe students with behavioral concerns.
- EBD and another disability specified in IDEA (e.g., learning disability, other health impairment, speech and language disorder), except for students with a dual diagnosis of moderate mental retardation or developmental disabilities (e.g., Cole, Davenport, Bambara, & Ager, 1997; O'Reilly et al., 2002) as these students typically participate in a functional skills curricula rather than traditional core curricula (Heckaman et al., 2000; Lane et al., 1999).
- A label of emotional disturbance (ED), as specified by IDEA (2004).
- Psychiatric diagnoses specified in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR; American Psychiatric Association, 2001) such as conduct disorder (CD) or oppositional defiant disorder (ODD).
- A general behavioral concern (e.g., noncompliance) and attention deficit/hyperactivity disorder, a group of students with attention and behavioral concerns that place them at heightened risk for behavior disorders.
- Psychiatric (e.g., ODD, CD) or educational (ED) diagnosis that co-occurred with an attention disorder (e.g., Ervin et al., 1998; Lane et al., 1999).

Second, all function-based interventions needed to take place in a secondary school setting, inclusive of middle, junior high, and high schools. If the study reported function-based interventions implemented in multiple school levels (e.g., elementary and middle), only results of the investigation taking place at the secondary school were included (e.g., DePaepe, Shores, Jack, & Denny, 1996; Gunter, Jack, Shores, Carrell, & Flowers, 1993; Stage et al., 2006). Interventions implemented in clinics, day treatment centers, diagnostic centers, or residential day treatment centers (e.g., Platt, Harris, & Clements, 1980) were excluded as the purpose of this review was to examine school-based interventions conducted in secondary schools. If the school level was not stated, the article was excluded unless the student was 13 years or older as there was very limited possibility that a 13-year-old would still be in elementary school.

Third, a functional assessment had to be conducted, yielding a hypothesis regarding the reason why the target behavior occurred. Functional assessment procedures—descriptive (e.g., interview, behavior rating scales, direct observation) or experimental (e.g., functional analysis)—must have been delineated. Consistent with other review articles, at least one of the preceding functional assessment procedures must have been employed in the methodological procedures and a hypothesis statement generated from the functional assessment results (Heckaman et al., 2000; Lane et al., 1999). Further, the article needed to include an intervention based on functional as-

sessment results (Heckaman et al.) and evaluated using single-case methodology. Articles that included only functional assessment results, functional analyses that did not lead to sustained interventions, or those with interventions not based on functional assessment results were excluded (e.g., DePaepe et al., 1996; Ervin et al., 2000).

Fourth, the studies must have reported a graphic display of student outcomes for individual students. Studies reporting only narrative outcomes (e.g., Sterling-Turner, Robinson, & Wilczynski, 2001) were excluded. We viewed this visual display as essential to evaluate the accuracy of treatment-outcome results and the analytical tools (e.g., stability, level, trend) employed. Further, studies reporting graphic display of group outcomes (e.g., Center, Deitz, & Kaufman, 1982) were excluded as they did not allow for inspection of individual outcomes.

Finally, only articles published in peer-reviewed journals were included in this review. Dissertations, book chapters, and monographs were excluded because our goal was to draw conclusions based on information that had withstood the peer review process.

Of the 33 articles identified in the initial search, 12 articles met the inclusion criteria as determined by all three authors. These articles were coded independently by the first and third authors as described in the following section.

Coding Procedures for Quality Indicators

Articles that met the inclusion criteria were read in their entirety by all three authors and coded by the first and third authors. Each article was coded along the 21 components constituting the seven quality indicators specified by Horner et al. (2005; see Table 1): (a) describing participants and settings; (b) dependent variable; (c) independent variable; (d) baseline; (e) experimental control/internal validity; (f) external validity; and (g) social validity. Specifically, each component was evaluated as being present or absent according to the guidelines in the sections that follow.

Describing Participants and Settings. Per Horner et al. (2005), this indicator contained three components: (a) participant description, (b) participant selection, and (c) setting description. To meet the first component, more than a general definition (e.g., EBD) was required. Participants had to be described in sufficient detail that included (a) the specific disability as well as (b) the method used to determine the disability. Participant selection criteria needed to be defined precisely enough to allow replication (e.g., quantifiable data to indicate replication selection criteria). Setting description required a description of the physical setting that also included sufficient details (e.g., number of adults present, room arrangement) that allowed others to recruit similar participants from similar settings. The coding reliability was as follows: participant description 83.33%; participant selection 100%; and setting description 91.67%.

Dependent Variable. Horner et al. (2005) identified five components to determine the quality of the dependent variables. First, the description of each dependent variable had to be operationally defined. If more than one dependent variable was reported, and both variables were not defined precisely, then this component was considered absent. Second, each dependent variable needed to be measured using a procedure that produced a quantifiable index such as the frequency of a given behavior per minute. Third, the measurement of the dependent variable needed to be valid and described with sufficient precision to allow for replication (e.g., appropriate system of measure dependent on the nature of the target behavior [whole interval for variables such as engagement, partial interval for variables such as disruption] with details of the data collection procedures provided).

Fourth, the dependent variable needed to be measured repeatedly over time. We further defined this component to require a minimum of 3 data points per condition (Kennedy, 2005). As mentioned by Kennedy, 3 data points per phase is an acceptable standard among researchers employing single-case methodology. Fifth, data needed to be reported regarding the reliability or interobserver agreement (IOA) of each dependent variable. Further, Horner et al. indicated that IOA levels had to meet the following minimum standards: IOA = 80% and Kappa = 60%. Because some articles reported ranges as well as means, we further defined this component to re-

TABLE 1

Application of Quality Indicators Posed by Horner et al. (2005)

Quality Indicator	*Schloss, Kane, & Miller (1981)*	*Knapczyk (1988)*	*Knapczyk (1992)*	*Gunter, Jack, Shores, Carrell, & Flowers (1993)*	*Ervin, DuPaul, Kern & Friman (1998)*	*Penno, Frank, & Wacher (2000)*
Describing participants and setting	No (0.00)	No (0.33)	No (0.33)	No (0.33)	No (0.67)	No (0.33)
Participant description	No (0.00)	No (0.00)	No (0.00)	No (0.00)	Yes (0.33)	No (0.00)
Participant selection	No (0.00)	No (0.00)	No (0.00)	No (0.00)	No (0.00)	No (0.00)
Setting description	No (0.00)	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)
Dependent variable	No (0.40)	No (0.80)	Yes (1.00)	Yes (1.00)	No (0.80)	Yes (1.00)
Description	No (0.00)	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)
Quantifiable measurement	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)
Valid and well-described measurement	No (0.00)	No (0.00)	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)
Measured repeatedly	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)	No (0.00)	Yes (0.20)
Interobserver agreement	No (0.00)	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)
Independent variable	No (0.33)	No (0.33)	No (0.33)	Yes (1.00)	Yes (1.00)	Yes (1.00)
Independent variable description	No (0.00)	No (0.00)	No (0.00)	Yes (0.33)	Yes (0.33)	Yes (0.33)
Systematically manipulated	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)
Fidelity of implementation	No (0.00)	No (0.00)	No (0.00)	Yes (0.33)	Yes (0.33)	Yes (0.33)
Baseline	No (0.50)	Yes (1.00)	Yes (1.00)	Yes (1.00)	No (0.50)	Yes (1.00)
Repeated measurement, established pattern	Yes (0.50)	Yes (0.50)	Yes (0.50)	Yes (0.50)	No (0.00)	Yes (0.50)
Description	No (0.00)	Yes (0.50)	Yes (0.50)	Yes (0.50)	Yes (0.50)	Yes (0.50)
Experimental control/internal validity	No (0.67)	No (0.67)	No (0.67)	Yes (1.00)	No (0%)	No (0%)
Three demonstrations of experimental effect	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)	No (0.00)	No (0.00)
Internal validity	No (0.00)	No (0.00)	No (0.00)	Yes (0.33)	No (0.00)	No (0.00)
Pattern of results	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)	No (0.00)	No (0.00)
External validity	No (0.00)	No (0.00)	No (0.00)	No (0.00)	No (0.00)	No (0.00)
Social validity	No (0.50)	No (0.75)	No (0.75)	No (0.75)	No (0.75)	No (0.50)
DV is socially important	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)
Change in DV is socially important	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)	No (0.00)
IV is practical and cost effective	No (0.00)	No (0.00)	No (0.00)	No (0.00)	No (0.00)	No (0.00)
Use in typical contexts	No (0.00)	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)
Number of indicators: Absolute coding	0	1	2	4	1	3
Number of indicators: Weighted coding	2.4	3.88	4.08	5.08	3.72	3.88

continues

TABLE 1 *(Continued)*

Quality Indicator	*Smith & Sugai (2000)*	*March & Horner (2002)*	*Hoff, Ervin, & Friman (2005)*	*Ingram, Lewis-Palmer, & Sugai (2005)*	*Liaupsin, Umbreit, Ferro, Urso, & Upreti (2006)*	*Stage, Jackson, Moscovitz, Erickson, Thurman, & Jessee, et. al. (2006)*
Describing participants and setting	Yes (1.00)	No (0.33)	No (0.67)	No (0.33)	No (0.67)	No (0.33)
Participant description	Yes (0.33)	No (0.00)	Yes (0.33)	No (0.00)	Yes (0.33)	No (0.00)
Participant selection	Yes (0.33)	Yes (0.33)	No (0.00)	Yes (0.33)	No (0.00)	Yes (0.33)
Setting description	Yes (0.33)	No (0.00)	Yes (0.33)	No (0.00)	Yes (0.33)	No (0.00)
Dependent variable	Yes (1.00)	No (0.80)	No (0.80)	No (0.60)	Yes (1.00)	No (0.40)
Description	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)	No (0.00)
Quantifiable measurement	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)
Valid and well-described measurement	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)	Yes (0.20)	No (0.00)
Measured repeatedly	Yes (0.20)	Yes (0.20)	No (0.00)	No (0.00)	Yes (0.20)	Yes (0.20)
Interobserver agreement	Yes (0.20)	No (0.00)	Yes (0.20)	No (0.00)	Yes (0.20)	No (0.00)
Independent variable	Yes (1.00)	No (0.67)	No (0.67)	Yes (1.00)	Yes (1.00)	No (0.67)
Independent variable description	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)
Systematically manipulated	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)	Yes (0.33)
Fidelity of implementation	Yes (0.33)	No (0.00)	No (0.00)	Yes (0.33)	Yes (0.33)	No (0.00)
Baseline	Yes (1.00)	Yes (1.00)	No (0.50)	No (0.50)	Yes (1.00)	No (0.50)
Repeated measurement, established pattern	Yes (0.50)	Yes (0.50)	No (0.00)	Yes (0.50)	Yes (0.50)	Yes (0.50)
Description	Yes (0.50)	Yes (0.50)	Yes (0.50)	No (0.00)	Yes (0.50)	No (0.00)
Experimental control/internal validity	Yes (1.00)	No (0.67)	No (0.00)	No (0.00)	No (0.00)	No (0.00)
Three demonstrations of experimental effect	Yes (0.33)	Yes (0.33)	No (0.00)	No (0.00)	No (0.00)	No (0.00)
Internal validity	Yes (0.33)	No (0.00)	No (0.00)	No (0.00)	No (0.00)	No (0.00)
Pattern of results	Yes (0.33)	Yes (0.33)	No (0.00)	No (0.00)	No (0.00)	No (0.00)
External validity	Yes (1.00)	No (0.00)	No (0.00)	No (0.00)	No (0.00)	No (0.00)
Social Validity	Yes (1.00)	No (0.75)	No (0.75)	No (0.75)	No (0.50)	No (0.50)
DV is socially important	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)
Change in DV is socially important	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)	No (0.00)	No (0.00)
IV is practical and cost effective	Yes (0.25)	No (0.00)	No (0.00)	No (0.00)	No (0.00)	No (0.00)
Use in typical contexts	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)	Yes (0.25)
Number of indicators: Absolute coding	7	1	0	1	3	1
Number of indicators: Weighted coding	7	4.22	3.39	3.18	4.17	2.4

Note. DV = dependent variable; IV = independent variable. Subscales do not always total to 1.00 due to rounding. Numbers in parentheses refer to the weighted value of the presence of each intervention component constituting an indicator, with each component contributing an equal proportion of the quality indicator.

quire only the means to be at or above these criteria. Namely, the component was considered present if the mean met criteria, regardless if the range of scores reported included values below the minimum criteria. If the article reported IOA and Kappa values (e.g., March & Horner, 2002), then both minimum criteria had to be met to be considered present. Finally, it was necessary for IOA or Kappa values to be presented for each measure and each phase. The coding reliability was as follows: dependent variable description 100%; quantifiable measurement 100%; valid and well-defined measurement 100%; measured repeatedly 100%; and IOA 91.67%.

Independent Variable. Horner et al. (2005) delineated three components as being necessary for single-subject studies to meet the independent variable quality indicator. First, the independent variable needed to be described precisely to allow for replication. This included documentation of required materials and explicit reporting of specific procedures. General descriptions (e.g., token economy) were considered insufficient to meet the expectation. Second, the independent variable needed to be systematically manipulated by the intervention agent (e.g., teacher, paraprofessional). Third, fidelity of implementation was considered "highly desirable" (Horner et al., p. 174). Horner et al. defined this as continuous direct measurement of the independent variable or a parallel form of assessment. To further define this component, we added that fidelity needed to be both measured explicitly and the data reported (Lane et al., 1999). The coding reliability was as follows: independent variable description 100%; systematically manipulated 100%; and fidelity of implementation 91.67%.

Baseline. Horner et al. (2005) indicated that baseline conditions needed to include repeated measurement, with an established pattern of responding that could be used to anticipate or predict future behavior in the absence of an intervention. They reported that baseline phases should include multiple data points. Specifically, Horner et al. stated the following: "five or more, although fewer data points are acceptable in specific cases" (Horner et al., p. 168). In addition, they required that either (a) a trend in the predicted direction of the intervention effect not be present or (b) that the trend be countertherapeutic. For this component, we established a minimum of 3 data points rather than 5, reasoning that 5 may be an unnecessarily high number. As Kennedy (2005) stated, "[A] baseline needs to be as long as necessary but no longer. The goal of baseline is to establish patterns of behavior to compare to intervention. Therefore, a baseline needs only be long enough to adequately sample this pattern" (p. 38). The second component necessary to establish the quality of baseline was that baseline conditions be described with sufficient detail to allow for replication. We clarified this indicator by establishing that the baseline description needed to include information on "who did what to whom, where were those actions taken, and when did those actions occur" (Lane, Worley, Reichow, & Rogers, 2006, p. 226). The coding reliability was as follows: repeated measurement/established pattern 100% and description 100%.

Experimental Control/Internal Validity. Horner et al. (2005) established experimental control/internal validity as being evident when

> the design documents three demonstrations of the experimental effect at three different points in time with a single participant (within-subject replication), or across different participants (inter-subject replication). An experimental effect is demonstrated when predicted change in the dependent variable covaries with manipulation of the independent variable. (p. 168)

They indicated that three components needed to be addressed to establish experimental control. First, the design must include at least three demonstrations of experimental effect at three different time points. In instances in which fewer than three demonstrations were documented in a given experiment, this component was considered absent. Second, the design needed to control for common threats to internal validity. In addition to requiring established designs that met this criteria (e.g., ABAB; BABA; changing criterion, multiple baseline with three legs, alternating treatment), we also required that treatment integrity be assessed and reported given that the absence of treatment integrity poses a severe threat to internal validity (Gresham, 1989).

Finally, Horner et al. (2005) required a pattern of responding that documented experimental control. They indicated that visual analysis techniques, which involve interpretation of level, trend, and variability of performance during each phase as well as other techniques (e.g., immediacy of effects, magnitude of change, percentage of overlapping data points, and consistency of data patterns) should be used to determine if this component was met. For our coding procedures, we determined that authors did not need to discuss each element (level, trend, variability) in text. If a graph with individual student-level data was displayed and the reader could examine level, trend, and variability and the graph suggested a functional relation between the introduction of the independent variable and corresponding changes in the dependent variables, the component was coded as present. The coding reliability was as follows: three demonstrations of experimental effect 100%; internal validity 91.67%; and pattern of results 100%.

External Validity. Horner et al. (2005) recommended documenting external validity by replicating experimental effects across participants, settings, behaviors, or materials. Consistent with Tankersley, Cook, and Cook (in press), we interpreted this quality indicator to require replication across one of the following: participants, setting, behavior, or materials. To further clarify criteria for external validity, we required studies to (a) include three replications in one of those categories as recommended by Horner et al., and (b) meet all three previously stated criteria for experimental control/internal validity to be considered as possibly having external validity given that internal validity is essential to establishing external validity (Wolery, 2007 personal communication). The coding reliability was as follows: 100% external validity.

Social Validity. Horner et al. (2005) identified social validity as the final quality indicator, which referred to the social significance of the goals, social acceptability of the treatment procedures, and the social importance of the effects (Baer, Wolf, & Risley, 1968). They identified four components for this indicator. First, the dependent variables needed to be socially valid. Second, the change in the dependent variable had to be socially important, defined as a "demonstration that the intervention produced an effect that met the defined, clinical need" (Horner et al., p. 172). We further defined this component as being present if (a) there was a measure of social validity and the evidence from that measure reported a socially meaningful change in the desired direction or (b) a functional relation was evident between the introduction of the independent variable and change in the target behavior (e.g., reduction in aggression).

Third, the independent variable was practical and cost effective. We clarified this component by stating that cost effectiveness must be stated explicitly. Practicality was defined as a study conducted in a typical setting with traditional intervention agents and materials typically found in the identified setting. Finally, use in typical contexts was defined as the following:

> demonstration that typical intervention agents (a) report procedures to be acceptable, (b) report the procedures to be feasible within available resources, (c) report the procedure to be effective, and (d) choose to continue use of the intervention procedures after formal support/expectation of use is removed. (Horner et al., 2005, p. 172)

We coded this component as present if any one of these four practices was reported. The coding reliability was as follows: social importance of the dependent variable 100%; change in dependent variable is socially important 75%; independent variable is practical and cost effective 100%; and used in typical contexts 100%.

An overarching framework when coding the quality indicators was to evaluate the studies based on what the researchers reported either in text or in visual display, and not in our interpretations. The modifications to the components constituting the quality indicators were developed to (a) refine definitions to increase consistency across raters and (b) allow more transparent criteria for the reader.

Evaluation Procedures for Determining Evidence-Based Practice Using Single-Subject Research

We then applied the five standards for an evidence-based practice proposed by Horner et al.

(2005) to the body of literature examining the effectiveness of function-based intervention conducted with secondary-age students with and at risk for EBD. The goal was to determine if single-subject research studies document this practice as evidence based. The five standards necessary to document a practice as evidence-based included the following:

1. The practice was defined operationally in text.
2. The authors defined the context and outcomes associated with the practice.
3. The practice was implemented with fidelity.
4. Findings document the introduction of the practice as functionally related to change in the dependent variables.
5. Experimental effects are replicated across sufficient number of peer-reviewed studies (n = 5) published in refereed journals, across three different researchers at three different geographical locales, and include at least 20 participants from five or more studies.

RESULTS

In the results section, we address the first two purposes of this review by answering the following questions: To what extent do the studies identified for inclusion meet the quality indicators posed by Horner et al. (2005)? To what extent do the studies addressing the quality indicators support function-based interventions for secondary-age students with EBD as an evidence-based practice?

STUDIES OF FUNCTION-BASED INTERVENTIONS FOR SECONDARY-AGE STUDENTS WITH EBD

Findings of a Field Test of Quality Indicators

Quality Indicator 1: Describing Participants and Setting. Results revealed that only 1 of the 12 studies reviewed met all three components (participant description, participant selection criteria, and setting description) constituting the quality indicator for describing participants and setting (Smith & Sugai, 2000; see Table 1). Three studies addressed two of the three components by reporting descriptions of the participant and setting that were precise enough to facilitate replication (Ervin et al., 1998; Hoff, Ervin, & Friman, 2005; Liaupsin, Umbreit, Ferro, Urso, & Upreti, 2006). However, despite the thorough description of the students' disabilities or condition and the procedures used to determine their disabilities, studies did not describe the process used to select participants with replicable precision. Of the 8 remaining studies, all but 1 (Schloss, Kane, & Miller, 1981) met at least one component constituting this quality indicator. Four studies reported the critical features of the setting, but these were not precise enough in describing the participants or the participant selection process so were not included (Gunter et al., 1993; Knapczyk, 1988, 1992; Penno, Frank, & Wacker, 2000). In contrast, the remaining three studies provided detailed descriptions of the participant selection process, but these did not provide enough detail in describing the participants or setting to allow for replication (Ingram, Lewis-Palmer, & Sugai, 2005; March & Horner, 2002; Stage et al., 2006). In some instances, the level of precision for describing the participant selection process was particularly detailed. For example, March and Horner stated the following:

> Three participants were selected based on (a) no decrease in their rate of discipline contacts following involvement with the BEP program, (b) documentation of at least five office discipline referrals during the first 4 months of the new academic year, (c) nomination by BEP team members, (d) student assent and parent consent. (p. 162)

Setting was the most frequently addressed, with 8 out of 12 studies meeting the coding criteria for this component. In contrast, only 4 studies described participants with sufficient detail to afford replication (Ervin et al., 1998; Hoff et al., 2005; Liaupsin et al., 2006; Smith & Sugai, 2000). Similarly, 4 studies described participant selection criteria with replicable precision (Ingram et al., 2005; March & Horner, 2002; Smith & Sugai; Stage et al., 2006).

Quality Indicator 2: Dependent Variables. Five studies met the quality indicator for dependent variables as evidenced by addressing the five components (description, quantifiable measurement, valid and well-described measurement, repeated measurement, and IOA) constituting this indicator (Gunter et al., 1993; Knapczyk, 1992; Liaupsin et al., 2006; Penno et al., 2000; Smith & Sugai, 2000). Four studies met coding criteria for all but one component (Ervin et al., 1998; Hoff et al., 2005; Knapcyzk, 1988; March & Horner, 2002); 1 study met criteria for three components (Ingram et al., 2005); and 2 studies met criteria for two components: quantifiable measurement and repeated measurement (Schloss et al., 1981; Stage et al., 2006).

In all studies, each dependent variable was measured in such a manner that produced a quantifiable index (e.g., percentage of intervals on-task; Ervin et al., 1998) and all but two studies (Schloss et al., 1981; Stage et al., 2006) operationally defined all dependent variables. In the latter study, all behavior codes were stated, but not all terms were operationally defined. The majority of studies (n = 9) reported a valid and well-described measurement system. For example, Liaupsin et al. (2006) described data collection of on-task behavior as follows "30-s whole interval recording procedure. Observations were 20 min in length and began 5 to 10 min after the assignment of independent class work or reading" (p. 584). In addition, nine studies measured the dependent variables repeatedly over time according to coding criteria (minimum of 3 data points per phase). In instances when this component was not met, there were typically fewer than 3 data points in a phase. For example, in the Ervin et al. (1998) study, one of the students, Joey, had just 1 datum point in the return to baseline phase. Finally, criteria for the IOA component (IOA > 80%; Kappa > 60%) were met in eight studies. However, in some cases IOA was reported as an overall mean, but not for each dependent variable individually (e.g., Ingram et al., 2005; Stage et al.). In other cases, the criterion for IOA criteria was met, yet the criterion for Kappa was not met (e.g., March & Horner, 2002).

Quality Indicator 3: Independent Variable (IV). Six studies met the quality indicator for independent variable as evidenced by addressing the three components (IV description, systematically manipulated, fidelity of implementation) constituting the quality indicator (Ervin et al., 1998; Gunter et al., 1993; Ingram et al., 2005; Liaupsin et al., 2006; Penno et al., 2000; Smith & Sugai, 2000). Three studies met two components: independent variable description and systematic manipulation of the independent variable (Hoff et al., 2005; March & Horner, 2002; Stage et al., 2006); yet, these studies did not address implementation fidelity. The final three studies addressed one out of three components, with all three studies systematically implementing the independent variable (Knapczyk, 1988, 1992; Schloss et al., 1981).

In all studies (n = 12) the independent variable was systematically manipulated by the experimenter; of these studies, 9 described the intervention procedures with replicable precision. Six studies measured and reported treatment fidelity. The 3 studies not meeting expectations for fidelity were published between 1981 and 1992 (Knapczyk, 1988, 1992; Schloss et al., 1981). However, it should be noted that the importance of treatment integrity was not emphasized in the literature until the 1980s as documented in articles written by Yeaton and Sechrest (1981) and Gresham (1989). March and Horner (2002) addressed the lack of treatment fidelity data as a limitation stating "a final limitation lies in the absence of treatment integrity data . . . the only process for documenting fidelity of procedural implementation was the weekly observation and feedback to teachers by the first author" (p. 168). Although Hoff et al. (2005) stated that "Kevin's teacher implemented all of the intervention strategies" (p. 50), they did not mention how (or if) they collected fidelity data. Finally, Stage et al. (2006) did monitor fidelity of data, but they reported poor fidelity of implementation (e.g., "In Gale's case, there was a complete lack of treatment fidelity within the general education setting." p. 468), thereby not meeting this component.

Quality Indicator 4: Baseline. Seven studies met the quality indicator for baseline as evidenced by addressing the two components (repeated measurement and established pattern description) constituting the quality indicator (Gunter et al., 1993; Knapczyk, 1988, 1992; Liaupsin et al., 2006; March & Horner, 2002; Penno et al.,

2000; Smith & Sugai, 2000). The remaining five studies met at least one of the two criteria for the baseline quality indicator. More specifically, three studies met at least one component, meeting expectations for an established pattern and repeated measurement (Ingram et al., 2005; Schloss et al., 1981; Stage et al., 2006). The other two studies met expectations for description of baseline conditions (Ervin et al., 1998; Hoff et al., 2005).

Ten studies met the criteria for reporting a baseline phase that included three or more data points and an established pattern of repeated measurement of a dependent variable that supported a patterned responding predictive of future behavior. However, 2 studies included fewer than the requisite number of data points in the return to baseline phase (Ervin et al., 1998; Hoff et al., 2005), although Hoff and colleagues acknowledge this as a "brief withdrawal of the intervention and return to baseline" (p. 51). Nine studies met the requisite criteria for describing the baseline condition. The remaining 3 studies did not describe the baseline condition precisely enough for replication (Ingram et al., 2005; Schloss et al., 1981; Stage et al., 2006).

Quality Indicator 5: Experimental Control/Internal Validity. Two studies (Gunter et al., 1993; Smith & Sugai, 2000) met the three components constituting this quality indicator: three demonstrations of experimental effect, internal validity, and pattern of results. Four studies met two components: three demonstrations of experimental effect and pattern of results, with internal validity not established (Knapczyk, 1988, 1992; March & Horner, 2002; Schloss et al., 1981). Six studies did not meet any of the components.

In terms of the components, six studies demonstrated experimental effect as evidenced by at least three demonstrations across participants (e.g., March & Horner, 2002; Schloss et al., 1981); setting (Knapczyk, 1988, 1992); or via an ABAB design (Gunter et al., 1993; Smith & Sugai, 2000). Based on coding criteria, experimental effect was scored as absent if there was an insufficient number of data points in a phase (e.g., Ervin et al., 1998; Hoff et al., 2005; Ingram et al., 2005) or if there were only two or fewer demonstrations evident (e.g., Liaupsin et al., 2006; Penno et al., 2000). Only two studies (Gunter et al.; Smith & Sugai) established internal validity according to the posed criteria. Several studies did not meet this component due to the absence of treatment integrity (e.g., Hoff et al.; Knapczyk, 1992; March & Horner; Schloss et al.; Stage et al.). Finally, six studies met the component of pattern of results that supported experimental control (Gunter et al.; Knapczyk, 1988, 1992; March & Horner; Schloss et al.; Smith & Sugai). The absence of sufficient data points in each phase prohibited studies from satisfying this component (e.g., Ervin et al.; Hoff et al.; Ingram et al.), as did the absence of sufficient demonstrations (e.g., Liaupsin et al.; Penno et al.; Stage et al.).

Quality Indicator 6: External Validity. Only one study established external validity according to the coding procedures (Smith & Sugai, 2000). In most studies, external validity was not established given that we defined the presence of internal validity as a prerequisite to external validity. Namely, the study needed to meet all components constituting the experimental control/internal validity indicator to have the possibility of experimental control. Thus, only two studies (Gunter et al., 1993; Smith & Sugai) had the possibility of meeting this indicator.

Quality Indicator 7: Social Validity. One study met the quality indicator for social validity as evidenced by addressing the four components (dependent variable is socially important, change in dependent variable is socially important, independent variable is practical and cost effective, and practice is used in typical contexts) constituting the indicator (Smith & Sugai, 2000). Seven studies met all components save for the third component, which required cost-effectiveness to be stated (Ervin et al., 1998; Gunter et al., 1993; Hoff et al., 2005; Ingram et al., 2005; Knapczyk, 1988, 1992; March & Horner, 2002). The remaining four studies met two of the four components, with three studies establishing the dependent variable as socially important and employing the independent variable in typical contexts (Liaupsin et al., 2006; Penno et al., 2000; Stage et al., 2006). The fourth study established the dependent variable as socially important and reported a change that was socially important (Schloss et al., 1981).

All studies established the dependent variable as socially important and 11 reported use of the

independent variable in typical contexts. Nine established the change in the dependent variable as socially important. Yet, only 1 study (Smith & Sugai, 2000) specifically stated that the intervention was both practical and cost effective, reporting that the intervention was "conducted in [an] actual classroom with minimal time or use of additional resources" (p. 215).

Function-Based Interventions for Secondary-Age Students With EBD: Determination of an Evidence-Based Practice

Given that only one study (Smith & Sugai, 2000) met all seven quality indicators, it is clear that function-based interventions conducted with secondary-age students with and at risk for EBD cannot yet be documented as an evidence-based practice according to Horner et al.'s (2005) standards. As a practice, function-based interventions involve (a) conducting descriptive and, in some cases, experimental tools to identify the function of the target behavior; (b) designing an intervention linked to functional assessment data to adjust antecedent conditions and to maintain consequences so that the student can acquire a more reliable, more efficient, functionally equivalent behavior; and (c) implementing the intervention with fidelity using an experimental design (e.g., multiple baseline, ABAB) that ensures experimental control. However, in the studies reviewed, the number of quality indicators met in entirety ranged from 0 to 7 (see Table 1). Moreover, only one study met four indicators (Gunter et al., 1993); two studies met three indicators (Liaupsin et al., 2006; Penno et al., 2000); one study met two indicators (Knapcyzk, 1992); and four studies met just one indicator (Ervin et al., 1998; Ingram et al., 2005; Knapczyk, 1988; March & Horner, 2002).

In addition, it should be noted that despite the specification of inclusion criteria, there was still variability in the functional assessment tools employed, student characteristics, and instructional setting. For example, although all studies reviewed met the inclusion criteria of having one functional assessment tool, a hypothesis, and an intervention linked to the functional assessment data, there still was variability in the functional assessment process used to identify the maintaining function of the target behavior (see Table 2). Some studies involved both teacher and student interviews (e.g., Ervin et al., 1998; Hoff et al., 2005; Ingram et al., 2005; Liaupsin et al., 2006; March & Horner, 2002; Penno et al., 2000; Smith & Sugai, 2000; Stage et al., 2006), yet other studies involved only teacher interviews. Likewise, several studies involved functional analyses of behavior (e.g., Ervin et al.; Hoff et al.; Penno et al.; Stage et al.). Second, the articles reviewed contained students with different facets of EBD as described in the article selection process. Finally, although all studies were conducted in school-based settings (e.g., self-contained schools, self-contained classrooms), and not in clinical settings, there was still heterogeneity in the settings. Thus, it should be noted that there was still variability in terms of target population, context, and functional assessment processes. Even if the results supported functional assessment-based interventions as an evidence-based practice for adolescents with or at risk for EBD according to quality indicators posed by Horner et al. (2005), the actual practice evaluated still may have contained variability in the components constituting the practice despite the inclusion criteria specified in this review.

DISCUSSION

Students with EBD pose significant challenges to parents, teachers, and society as a whole (Kauffman, 2005). Function-based interventions are one tertiary level, ideographic approach employed to meet the multiple needs of this population, particularly for elementary-age students (Lane et al., 1999). However, function-based interventions have not yet been established as an evidence-based practice for secondary-age students with EBD according to the criteria specified by Horner et al. (2005). This is unfortunate given that function-based interventions are mandated per IDEA for students with specific disciplinary circumstances (Kern et al., 2004)

In this analysis, a systematic literature review identified 12 studies of function-based interventions conducted with middle and high school students with and at risk for EBD in school settings.

TABLE 2

Functional Assessment Components

Functional Assessment Component	*Schloss, Kane, & Miller (1981)*	*Knapczyk (1988)*	*Knapczyk (1992)*	*Gunter, Jack, Shores, Carrell, & Flowers (1993)*	*Ervin, DuPaul, Kern, & Friman (1998)*	*Penno Frank, & Wacker (2000)*
Direct observations	no	yes	yes	yes	yes	yes
Teacher interview	yes	yes	yes	no	yes	yes
Student interview	yes	no	no	no	yes	yes
Parent interview	yes	no	yes	no	no	no
Other interview	no	no	no	no	no	no
Rating scales	no	no	no	yes	yes	no
Record search	no	no	no	no	no	yes
Functional analysis	no	no	no	no	yes	yes
Hypothesis statement	yes	yes	yes	yes	yes	yes
Intervention linked to assessment data	yes	yes	yes	yes	yes	yes

Functional Assessment Component	*Smith & Sugai (2000)*	*March & Horner (2002)*	*Hoff, Ervin, & Friman (2005)*	*Ingram, Lewis-Palmer, & Sugai (2005)*	*Liaupsin Umbriet, Ferro, Urso, & Upreti (2006)*	*Stage, Jackson, Moscovitz, Erickson, Thurman, & Jessee, et al. (2006)*
Direct observations	yes	yes	yes	yes	yes	yes
Teacher interview	yes	yes	yes	yes	yes	yes
Student interview	yes	yes	yes	yes	yes	yes
Parent interview	no	no	no	no	no	yes
Other interview	no	no	no	no	no	no
Rating scales	no	no	yes	no	no	yes
Record search	yes	yes	no	no	yes	no
Functional analysis	no	no	yes	no	no	yes
Hypothesis statement	yes	yes	yes	yes	yes	yes
Intervention linked to assessment data	yes	yes	yes	yes	yes	yes

Application of the core quality indicators for single-subject research revealed only one study (Smith & Sugai, 2000) as meeting all 21 components constituting the seven quality indicators posed by Horner and colleagues (2005). Given that only one study met this rigorous set of indicators, there is an insufficient number of studies conducted that meet the requisite standards for qualifying a practice as "evidence-based" according to the criteria set forth by Horner and colleagues. However, we contend that this assessment may be based on indicators that may be somewhat too rigorous. In the sections that follow we (a) offer illustrations of how some of the indicators may exceed reasonable standards and (b) propose a different approach to evaluating a given study against the posed quality indicators.

Quality Indicators: Reasonable Standards?

As we coded the articles in the review, we discussed certain components that may be so stringent that they excluded studies that do, in fact, make a meaningful contribution to the knowledge base. Specifically, we felt that the require-

ments for describing participants, establishing repeated measurement of the dependent variable, repeated measurement and established pattern for baseline, and stating cost-effectiveness as a component of the social validity indicator may need to be reconsidered.

In this analysis, a systematic literature review identified 12 studies of function-based interventions conducted with middle and high school students with and at risk for EBD in school settings.

Describing Participants. For example, in Quality Indicator 1: Describing Participants and Settings, the first component focused on participant description. To meet requisite criteria for this component, the authors needed to report the specific disability or condition and the "specific instrument and process used to determine their disability" (Horner et al., 2005, p. 167). It may be that the latter component is beyond reasonable at this time. Although it is important to ensure precision for purposes of replication, it may be more reasonable to require that the process (e.g., as determined by a multidisciplinary team) be acceptable rather than requiring specific instruments. This is particularly true given information available in cumulative files and space limitations associated with publication efforts.

Repeated Measurement. As part of Quality Indicator 2: Dependent Variable, component four required that dependent variables be measured repeatedly over time, and Quality Indicator 4: Baseline established the need for 5 data points in baseline, which we altered to require a minimum of 3 data points. Yet, according to Kennedy (2005), the "goal of baseline is to establish patterns of behavior to compare to intervention. Therefore, a baseline needs only be long enough to adequately sample this pattern" (p. 38). Consider the study by Ervin et al. (1998) in which fewer than 3 data points were collected during the reversal phases. One could argue that because of the dramatic change in level, additional data points were not warranted in the return to baseline phase. However, for purposes of this review, all articles with fewer than 3 data points were reported as not meeting the component of repeated measurement.

Failure to meet the requisite number of data points per phase also influenced the extent to which the internal validity indicator was met. Again looking at the Ervin et al. (1998) study, the return to baseline phase for Joey had but 1 datum point, which did not meet criteria for baseline requirement. Thus, this study did not meet the internal validity criteria. Because internal validity is required to establish external validity (Wolery, 2007 personal communication), this also precluded this study from meeting the external validity components. Yet, despite the limited number of data points, the argument could be made for experimental control given the clear changes in level.

These same ramifications were recognized when coding the study conducted by Hoff et al. (2005). The brief return to baseline (2 data points) did not meet our minimum criteria of 3 data points per phase. Therefore, the article was coded as not having at least three demonstrations of experimental effect and the pattern of results was not sufficient given that only 2 data points were in the return to baseline condition. Because internal validity was not established, external validity was absent as well according to our coding procedures. However, in inspecting the graph, there was a very clear change in level and possibly trend when the intervention was withdrawn. This serves as another illustration as to the possibility that some of the components defining each quality indicator (e.g., requirement of a minimum of 3 data points) may be too stringent. Thus, some studies may be excluded that do lend support for a given practice.

Cost Effectiveness. The third component of Quality Indicator 7: Social Validity required that the intervention be "practical and cost effective" (Horner et al., 2005, p. 174). Concordant with Tankersley, Cook, and Cook's (in press) efforts to evaluate Horner et al.'s quality indicators based on information reported in text, our coding system required the cost-effectiveness of an intervention to be stated explicitly. Yet this requirement may be too rigorous because only one study (Smith & Sugai, 2000) explicitly mentioned cost-effectiveness of the intervention. Moving forward, it may be wise to offer clarifying points for evalu-

ating cost-effectiveness as many studies may indeed be cost-effective in the sense that the benefits outweigh the costs (e.g., time, resources), even though cost-effectiveness is not computed or discussed explicitly. One could argue that a practice's cost-effectiveness could be assessed indirectly by looking at social validity or treatment integrity data. Namely, if the intervention was too costly in terms of time or resources, then it would be apt to receive a negative social validity rating or be implemented with low fidelity (Lane & Beebe-Frankenberger, 2004)

We do recognize that it is difficult to develop indicators and coding practices that can successfully capture the contribution and qualities of all studies. For example, the Penno et al. (2000) study did not meet our criteria for establishing a socially important change in the dependent variable. However, it should be noted that the authors reported "of particular importance is the finding that behavior problems were reduced for 2 of three participants even though the instructional modifications were designed to enhance academic performance" (Penno et al., p. 341). The coding system we applied overlooked this finding. Further, as we evaluated studies that were published more than 2 decades ago, it is important to note that standards for research shift over time. For example, the three studies not mentioning or reporting fidelity of the independent variable, required as part of Quality Indicator 3: Independent Variable, were published between 1981 and 1992 (Knapczyk, 1988, 1992; Schloss et al., 1981)—prior to the emphasis placed on treatment integrity. Finally, the study by Stage et al. (2006) reported three cases, but article selection procedures restricted coding to only secondary-age students. Consequently, the two other applications to younger students—which met many of the indicators—were not reported in this review.

Application of the Indicators: A Modified Approach

Rather than evaluating only those studies that met all indicators in entirety, another approach might be to impose an 80% minimum criteria with "credit" or recognition of the components that were addressed in a given quality indicator. For example, the dependent variable quality indicator contains five components that need to be addressed. Moving forward, we may want to consider weighting each component, with each component contributing an equal proportion of the quality indicator. In the case of the dependent variable quality indicator, each component would be weighted as contributing to 20% of the total score for the indicator. To illustrate, consider the article by Ervin and colleagues (1998). This study met the requirements for description, quantifiable measurement, valid and well-described measurement, and IOA. Yet it did not meet the requirements for measured repeatedly. Rather than scoring this indicator as a zero for omitting one of the five components, a weighted scoring could be employed as follows:

> DV quality indicator = ((description)(1) (.20)) + ((quantifiable measurement)(1) (.20)) + ((valid and well-described measurement)(1) (.20)) + ((measured repeatedly)(0) (.20)) + ((IOA)(1)(.20))

In this case, rather than applying an absolute coding system of "met" or "not met," the study could receive "partial credit" for the components that were addressed. In the above illustration of the dependent variable indicator, the study would receive an overall score of .80 rather than receiving a zero.

If this method was applied to all indicators for this study, then the overall quality indicator composite score for the Ervin et al. (1998) study with partial credit would be 3.72 (describing participants = 0.67; dependent variable = 0.80; independent variable = 1.00; baseline = 0.50; experimental control = 0.00; external validity = 0.00; social validity = .75) as opposed to the current score of meeting one out of seven indicators. Such a scoring system would reveal a more precise, detailed description of the critical components addressed in the study.

In Table 2, we present a total score for each article when scored using the presence or absence of each indicator as well as the partial-credit scoring system explained above. If we set a goal of studies achieving 80% of the indicators (80% x 7 indicators), then studies with a total score of 5.60 could be considered rigorous enough to be evaluated in the decision of whether or not a practice is

evidence-based. In this review, no additional studies would have been included for evaluating the evidence base. However, it is possible that such a coding procedure could influence the number of studies included in other literature reviews.

Yet another consideration would be to differentiate between the value of each indicator. Namely, are certain quality indicators (e.g., internal and external validity) more important than other indicators (e.g., social validity)? Some may argue that violating internal validity is a more serious concern than omitting social validity. If so, should the weighted value of each indicator or each component within each indicator be considered? Also, should the value of the indicator be dependent on the type of study (efficacy or effectiveness) being conducted? As we move toward conducting studies in more applied settings with less university support, should the value of certain indicators be viewed as more or less necessary?

Conclusion Comments: Considerations and Future Directions

As we conclude the task of applying the quality indicators and standards posed by Horner et al. (2005) to function-based interventions with secondary-age students with EBD, we offer the following comments. First, we applaud Horner and colleagues for the effort placed into developing quality indicators for single-case research. This was clearly a formidable—and necessary—task that will continue to influence how research proposals and subsequent investigations will be conducted. We value the concept of setting standards and hope that our goal of offering input as to where these indicators may be too stringent and in need of modification is received in the spirit intended: to establish scientifically valid, yet reasonable indicators for evaluating single-subject work.

Finally, in this field testing of the proposed quality indicators and standards for evidence-based practice, we want to point out that all articles met inclusion criteria of having employed at least one functional assessment procedure, stating a hypothesis, and linking the intervention to the assessment results. Yet, there was still variability in the functional assessment process. Namely, some studies included interviews from teachers and students (e.g., Penno et al., 2000); some included functional analyses (e.g., Hoff et al., 2005); and some included record searches (e.g., Liaupsin et al., 2006). Thus, although the interventions derived from functional assessment data were evaluated in terms of the quality indicators, the functional assessment process was not standardized (Kern et al., 2004; Sasso et al., 2001). We recommend that future reviews be considered in which a particular method of conducting function-based interventions, such as the model posed by Umbreit et al. (2007) be evaluated to determine if the specific model is an evidence-based practice.

Despite these considerations, this article offers an initial application of the core quality indicators and standards for evidence-based practices proposed by Horner et al. (2005) for single-case methodology to functional assessment-based interventions conducted with secondary-age students with EBD or at risk for developing EBD. Findings suggest that when assessed using the criteria proposed, this practice cannot be considered an evidence-based practice at this time. However, we contend that this practice holds promise. Certainly, additional high-quality research may result in the practice being considered evidence-based for the target population using these or similar standards. Weighting the criteria, assigning partial credit, or weighting indicators depending on the focus of the study may also be possible directions for refining the application of indicators; in this way, researchers are certain to include all meaningful and trustworthy studies of the practices and ensure that important contributions to this body of literature are not eliminated based on criteria being unattainable. In the years to come, it will be important to be thoughtful and careful as scholars and stakeholders use the proposed indicators. There is a delicate balance between maintaining high scientific rigor and potentially eliminating or ruling out the use of promising practices, such as function-based interventions for adolescents with and at risk for EBD, that are associated with improved behavioral and academic performance.

REFERENCES

References marked with an asterisk indicate experiments included in the analysis.

American Psychiatric Association. (2001). *Diagnostic and statistical manual of mental disorders* (DSM-IV-TR; 5th ed.). Washington, DC: Author.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Bullis, M., & Yovanoff, P. (2006). Community employment experiences of formerly incarcerated youth. *Journal of Emotional and Behavioral Disorders, 14,* 71–85.

Center, D. B., Deitz, S. M., & Kaufman, M. E. (1982). Student ability, task difficulty, and inappropriate classroom behavior. *Behavior Modification, 6,* 355–374.

Cole, C. L., Davenport, T. A., Bambara, L. M., & Ager, C. L. (1997). Effects of choice and task preference on the work performance of students with behavior problems. *Behavioral Disorders, 22,* 65–74.

Conroy, M. A., Dunlap, G., Clarke, S., & Alter, P. J. (2005). A descriptive analysis of positive behavioral intervention research with young children with challenging behavior. *Topics in Early Childhood Special Education, 25,* 157–166.

*DePaepe, P. A., Shores, R. E., Jack, S. L., & Denny, R. K. (1996). Effects of task difficulty on the disruptive on-task behavior of students with severe behavior disorders. *Behavioral Disorders, 21,* 216–225.

Drasgow, E., & Yell, M. L. (2001). Functional behavior assessment: Legal requirements and challenges. *School Psychology Review, 30,* 239–251.

Dunlap, G., & Childs, K. E. (1996). Intervention research in emotional and behavioral disorders: An analysis of studies from 1980–1993. *Behavioral Disorders, 21,* 125–136.

*Ervin, R. A., DuPaul, G. J., Kern, L., & Friman, P. C. (1998). Classroom based functional and adjunctive assessments: Proactive approaches to intervention selection for adolescents with attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis, 31,* 65–78.

*Ervin, R. A., Kern, L., Clarke, S., DuPaul, G. J., Dunlap, G., & Friman, P. C. (2000). Evaluating assessment-based intervention strategies for students with ADHD and comorbid disorders within the natural classroom context. *Behavioral Disorders, 25,* 344–358.

Fox, J., Conroy, M., & Heckaman, K. (1998). Research in functional assessment of the challenging behaviors of students with emotional and behavioral disorders. *Behavioral Disorders, 24,* 26–33.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71,* 149–164.

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18,* 37–50.

Gresham, F. M. (2002). Social skills assessment and instruction for students with emotional and behavioral disorders. In K. L. Lane, F. M. Gresham, & T. E. O'Shaughnessy (Eds.), *Interventions for children with or at risk for emotional and behavioral disorders* (pp. 242–258). Boston: Allyn & Bacon.

Gresham, F. M. (2004). Current status and future directions of school-based behavioral interventions. *School Psychology Review, 33,* 326–343.

*Gunter, P. L., Jack, S. L., Shores, R. E., Carrell, D. E., & Flowers, J. (1993). Lag sequential analysis as a tool for functional analysis of student disruptive behavior in classrooms. *Journal of Emotional and Behavioral Disorders, 1,* 138–148.

Heckaman, K., Conroy, M., Fox, J., & Chait, A. (2000). Functional assessment-based intervention research on students with or at risk for emotional and behavioral disorders in school settings. *Behavioral Disorders, 25,* 196–210.

*Hoff, K. E., Ervin, R. A., & Friman, P. C. (2005). Refining functional behavioral assessment: Analyzing the separate and combined effects of hypothesized controlling variables during ongoing classroom routines. *School Psychology Review, 34,* 45–57.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71,* 165–179.

Horner, R. H., & Sugai, G. (2000). School-wide behavior support: An emerging initiative. *Journal of Positive Behavior Interventions, 2,* 231–232.

Individuals With Disabilities Education Act Amendments of 1997. Pub. L. No. 105-17, Section 20, 111 Stat. 37 (1997). Washington, DC: U.S. Government Printing Office.

Individuals With Disabilities Education Improvement Act of 2004, 20 U.S.C. §§ 1400 et esq. (2004).

*Ingram, K., Lewis-Palmer, T., & Sugai, G. (2005). Function based intervention planning: Comparing the effectiveness of FBA function-based and non-function

based intervention planning. *Journal of Positive Behavior Interventions, 7,* 224–236.

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K., & Richman, G. S. (1982). Toward a functional analysis of self-injury. *Analysis and Intervention in Developmental Disabilities, 2,* 3–20. (Reprinted in *Journal of Applied Behavior Analysis, 29,* 133–135)

Kauffman, J. (2005). *Characteristics of emotional and behavioral disorders of children and youth* (8th ed.). Upper Saddle River, NY: Pearson Merrill Prentice Hall.

Kennedy, C. H. (2005). *Single-case designs for educational research.* Boston: Allyn & Bacon.

Kern, L., Childs, K., Dunlap, G., Clarke, S., & Falk, G. (1994). Using assessment-based curricular intervention to improve the classroom behavior of a student with emotional and behavioral challenges. *Journal of Applied Behavior Analysis, 27,* 7–19.

Kern, L., Delaney, B., Clarke, S., Dunlap, G., & Childs, K. (2001). Improving the classroom behavior of students with emotional and behavioral disorders using individualized curricular modifications. *Journal of Emotional and Behavioral Disorders, 9,* 239–247.

Kern, L., Hilt, A. M., & Gresham, F. (2004). An evaluation of the functional behavioral assessment process used with students with or at risk for emotional and behavioral disorders. *Education and Treatment of Children, 27,* 440–452.

*Knapczyk, D. R. (1988). Reducing aggressive behaviors in special and regular class settings by training alternative social responses. *Behavioral Disorders, 14,* 27–39.

*Knapczyk, D. R. (1992). Effects of developing alternative responses on the aggressive behavior of adolescents. *Behavioral Disorders, 17,* 247–263.

Lane, K. L., & Beebe-Frankenberger, M. E. (2004). *School-based interventions: The tools you need to succeed.* Boston: Allyn & Bacon.

Lane, K. L., Umbreit, J., & Beebe-Frankenberger, M. (1999). A review of functional assessment research with students with or at-risk for emotional and behavioral disorders. *Journal of Positive Behavioral Interventions, 1,* 101–111.

Lane, K. L., Wolery, M., Reichow, B., & Rogers, L. (2006). Describing baseline conditions: Suggestions for study reports. *Journal of Behavioral Education, 16,* 224–234.

*Liaupsin, C. J., Umbreit, J., Ferro, J. B., Urso, A., & Upreti, G. (2006). Improving academic engagement through systematic, function-based intervention. *Education and Treatment of Children, 29,* 573–591.

Loeber, R., Green, S. M., Lahey, B. B., Frick, P. J., & McBurnett, K. (2000). Findings on disruptive behavior disorders from the first decade of the developmental trend study. *Clinical Child and Family Psychology Review, 3,* 37–59.

*March, R. E., & Horner, R. H. (2002). Feasibility and contributions of functional behavioral assessment in schools. *Journal of Emotional and Behavioral Disorders, 10,* 158–170.

Morris, R. J., Shah, K., & Morris, Y. P. (2002). Internalizing behavior disorders. In K. L. Lane, F. M. Gresham, & T. E. O'Shaughnessy (Eds.). *Interventions for children with or at risk for emotional and behavioral disorders* (pp. 223–241). Boston: Allyn and Bacon.

Morrison, G. M., Robertson, L., Laurie, B., & Kelly, J. (2002). Protective factors related to antisocial behavior trajectories. *Journal of Clinical Psychology, 58,* 277–290.

Nelson, J. R., Benner, G. J., Lane, K., & Smith, B. W. (2004). An investigation of the academic and behavioral disorders in public school settings. *Exceptional Children, 71,* 59–74.

O'Reilly, M., Tiernan, R., Lancioni, G., Lacey, C., Hillery, J., & Gardiner, M. (2002). Use of a self-monitoring and delayed feedback to increase on-task behavior in a post-institutionalized child within regular classroom settings. *Education and Treatment of Children, 25,* 91–102.

*Penno, D. A., Frank, A. R., & Wacker, D. P. (2000). Instructional accommodations for adolescent students with severe emotional or behavioral disorders. *Behavioral Disorders, 25,* 325–343.

Platt, J. S., Harris, J. W., & Clements, J. E. (1980). The effects of individually designed reinforcement schedules on attending and academic performance with behaviorally disordered adolescents. *Behavioral Disorders, 5,* 197–205.

Quinn, M. M., Gable, R. A., Fox, J., Rutherford, R. B., Jr., Van Acker, R., & Conroy, M. (2001). Putting quality functional assessment into practice in schools: A research agenda on behalf of E/BD students. *Education and Treatment of Children, 24,* 261–275.

Sasso, G. M., Conroy, M. A., Stichter, J. P., & Fox, J. J. (2001). Slowing down the bandwagon: The misapplication of functional assessment for students with emotional or behavioral disorders. *Behavioral Disorders, 26,* 282–296.

Sasso, G. M., Reimers, T. M., Cooper, L. J., & Wacker, D. P. (1992). Use of descriptive and experimental analyses to identify the functional properties of aberrant behavior in school settings. *Journal of Applied Behavior Analysis, 25,* 809–821.

*Schloss, P. J., Kane, M. S., & Miller, S. (1981). Truancy intervention with behavior disordered adolescents. *Behavioral Disorders, 6,* 175–179.

*Smith, B. W., & Sugai, G. (2000). A self-management functional assessment-based behavior support plan for a middle school student with EBD. *Journal of Positive Behavior Interventions, 2,* 208–217.

*Stage, S. A., Jackson, H. G., Moscovitz, K., Erickson, M. J., Thurman, S. O., Jessee, W., & Olson, E. M. (2006). Using multimethod-multisource functional behavioral assessment for students with behavioral disabilities. *School Psychology Review, 35,* 451–471.

Sterling-Turner, H. E., Robinson, S. L., & Wilczynski, S.M. (2001). Functional assessment of distracting and disruptive behaviors in the school setting. *School Psychology Review, 30,* 211–226.

Tankersley, M., Cook, B. G., & Cook, L. (in press). A preliminary examination to identify the presence of quality indicators in single-subject research. *Education and Treatment of Children, 31.*

Umbreit, J., Ferro, J., Liaupsin, C., & Lane, K. (2007). *Functional behavioral assessment and function-based intervention: An effective, practical approach.* Upper Saddle River, N J: Prentice-Hall.

Umbreit, J., Lane, K. L., & Dejud, C. (2004). Improving classroom behavior by modifying task difficulty: Effects of increasing the difficulty of too-easy tasks. *Journal of Positive Behavior Interventions, 6,* 13–20.

Walker, H. M., Ramsey, E., & Gresham, F. M. (2004). *Antisocial behavior in school: Evidence-based practices* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Yeaton, W., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49,* 156–167.

ABOUT THE AUTHORS

KATHLEEN LYNNE LANE (CEC TN Federation), Associate Professor, Department of Special Education; **JEMMA ROBERTSON KALBERG,** Graduate of the Department of Special Education at Peabody College of Vanderbilt University; and **JENNA COURTNEY SHEPCARO** (CEC TN Federation), Graduate of the Department of Special Education at Peabody College of Vanderbilt University, Nashville, Tennessee.

We thank Robert Horner, Edward Carr, James Halle, Gail McGee, Samuel Odom, and Mark Wolery for their seminal article, *The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education.* Further, we thank Jessica Weisenbach and M. Annette Little for their contribution to the literature search.

Address correspondence to Kathleen Lynne Lane, Department of Special Education, Peabody College, Vanderbilt University, Nashville, TN 37203, (615) 322-8179 (e-mail: kathleen.lane@vanderbilt.edu).

Manuscript received August 2007; accepted March 2008.