Article

Hammill Institute on Disabilities

Journal of Emotional and Behavioral Disorders
2014, Vol. 22(3) 171–189

DOI: 10.1177/1063426613476092
jebd.sagepub.com
SAGE

A Review of Tier 2 Interventions Conducted Within Multitiered Models of Behavioral Prevention

Allison Leigh Bruhn, PhD[1], Kathleen Lynne Lane, PhD[2], and Shanna Eisner Hirsch, MEd[3]

Abstract

To support students' academic, behavioral, and social needs, many schools have adopted multitiered models of prevention. Because Tier 3 interventions are costly in terms of time and resources, schools must find efficient and effective Tier 2 interventions prior to providing such intense supports. In this article, we review the literature base on Tier 2 interventions conducted within the context of multitiered prevention models evidencing a Tier 1 behavioral plan. Article selection and coding procedures are described and results are presented. Finally, we summarize our findings of four research questions, reflect on limitations, and offer suggestions for future inquiry.

Keywords

Tier 2, targeted, secondary, intervention, behavior

Multitiered models of prevention such as positive behavior interventions and supports (PBIS) and response-to-intervention (RtI) have been adopted by schools across the country to help meet the needs of *all* students. In these models, data are used to determine the level of support for each student. The support continuum follows a three-tiered approach, including Tier 1 (i.e., universal), Tier 2 (i.e., targeted), and Tier 3 (i.e., intensive) levels. As the school-site team monitors the progress and outcomes of the Tier 1 plan using academic and/or behavioral data, students are identified for more targeted supports (e.g., small group instruction, check-in/check-out [CICO] procedures). The most intensive supports (Tier 3) interventions (e.g., functional assessment-based interventions, one-on-one tutoring) are reserved for students exhibiting the most resistance to Tier 1 and Tier 2 efforts. As students' needs change over time, so does the level of support provided. However, students receiving Tier 2 and Tier 3 supports continue to receive access to the Tier 1 prevention plan, as this is the foundation of Tier 1 prevention—delivery to *all* students. In light of the resource-intensiveness of Tier 3 interventions, one goal is to reduce the number of students requiring Tier 3 support. A possible method for achieving this aim is to first place students in a more efficient Tier 2 intervention. Because 10% to 15% of the school population may need Tier 2 support, these types of interventions must be efficient, practical, and feasible to implement.

A Closer Look at Tier 2 Prevention Efforts

Recently, Mitchell, Stormont, and Gage (2011) published the first review of Tier 2 interventions conducted within the context of multitiered, behavioral prevention models. The review yielded several important findings about Tier 1 implementation, Tier 2 intervention outcomes, and social validity. Results showed (a) a majority of studies did not report Tier 1 treatment integrity, (b) Tier 2 interventions had promise for improving behavior, and (c) interventions were found to be important, practical, and feasible (Mitchell et al., 2011).

Because the Mitchell et al. (2011) review did not include studies reporting only academic outcomes, a subset of studies examining the effects of Tier 2 interventions on the academic skills of students with behavior problems were excluded. Given many students who exhibit persistent behavior problems also may struggle academically (Kauffman & Landrum,

[1]The University of Iowa, USA
[2]The University of Kansas, Lawrence, USA
[3]The University of Virginia, Chartlottesville, USA

Corresponding Author:
Allison Leigh Bruhn, Department of Teaching and Learning, The University of Iowa, Iowa City, IA 52242, USA.
Email: Allison-bruhn@uiowa.edu

2009), we cannot negate the importance of academic interventions for this at-risk population. Accordingly, we extend the work of Mitchell and colleagues by expanding our inclusion criteria for studies conducted within multitiered prevention models to also include studies reporting only academic outcomes.

Another important aspect of the Mitchell et al.'s (2011) review included examining participant characteristics (i.e., gender, age, school level), as well as how participants were identified for Tier 2 support (Mitchell et al., 2011). Because replication of studies and generalization of effects is dependent on a thorough understanding of participants and setting (Horner et al., 2005), we extend our description to include ethnicity, student status (e.g., disability, screening, mental health), socioeconomic status, grade level, and geographic locale.

Finally, the review by Mitchell and colleagues spanned 8 years of research (i.e., 2002–2009). In the last 2 years (i.e., 2010–2011), additional treatment-outcome studies have been published focusing on Tier 2 interventions. To account for this increase in publications, we extend the previous review by including 2 additional years of publications and examine how the literature base has changed over time in terms of reporting Tier 1 treatment integrity.

The objectives of this review are to examine (a) the extent to which the Tier 1 behavioral plan was implemented prior to intervening with Tier 2 support, as well as how reporting this has changed over time; (b) participant identification, characteristics, and setting; (c) measurement issues and research design; and (d) intervention components and associated student outcomes.

Method

To determine the extent to which Tier 2 interventions have been implemented and evaluated within the context of multitiered models evidencing a Tier 1 behavioral plan, we conducted a comprehensive, systematic literature review. Articles were located using electronic, hand, and ancestral searches. Once articles were retrieved, they were read to determine whether they met inclusion criteria. Finally, articles meeting inclusion criteria were coded for the presence, absence, and description of various components (see Table 1).

Article Selection Procedures

Search procedures. First, an electronic search of articles published through 2011 was conducted using the databases *ERIC* (Education Resources Information Center) and *PsycInfo*. Search terms included the following: (a) positive behavior support, effective behavior support, three-tiered model, multitiered model, or multileveled model; (b) Tier 2, targeted, secondary; and (c) support or intervention. Each title and abstract was read to determine whether the article merited a full reading for inclusion criteria. For example, if the title or abstract indicated the article was about an experimental study examining the effects of a school-based intervention, then the article was read in entirety. Conversely, if that criterion was contraindicated (e.g., the article was a literature review, the study was not school-based) in the title or abstract, then the article was not read in entirety, and therefore not eligible for inclusion. Of the 483 articles yielded in the search, 35 articles were retained.

Next, the OSEP Technical Assistance Center on Positive Behavioral Supports and Interventions website (www.PBIS.org) was examined. Each article listed under *secondary intervention research* and appearing in a peer-reviewed journal was retrieved for further review, resulting in 16 new articles. Ancestral searches of reference lists of articles obtained through ERIC, PsycInfo, and www.PBIS.org also were conducted yielding 7 articles.

Finally, hand searches of the six journals featuring two or more retained articles from the electronic and ancestral searches were conducted (2002–2011), including: *Education and Treatment of Children, Journal of Positive Behavior Interventions, Journal of Special Education, Remedial and Special Education, Journal of Emotional and Behavioral Disorders, and Behavioral Disorders*. This yielded nine new articles.

Inclusion and exclusion criteria. From the electronic, website, ancestral, and hand searches, 67 articles were read to determine whether they met the following inclusion criteria. First, the study had to take place within a schoolwide, multitiered prevention model with evidence of a Tier 1 behavioral plan. Tier 2 interventions taking place outside the context of a tiered, schoolwide model were excluded (e.g., Chafouleas, Riley-Tillman, Sassu, LaFrance, & Patwa, 2007). Although some studies mentioned behavior plans at the classroom level, the use of a schoolwide, multitiered model was not clear, and thus these studies were excluded (e.g., Cartledge, Sentell, Loe, Lambert, & Reed, 2001). If the presence of a multitiered model with a Tier 1 behavioral plan was mentioned to provide a context for the study, but explicitly stated the multitiered model was not the focus of the article, it was excluded (e.g., Kamps & Greenwood, 2005). At a minimum, the article had to briefly describe the presence of a universal, schoolwide, primary, or Tier 1 plan (e.g., Gresham, Van, & Cook, 2006).

Second, the article had to include a Tier 2 intervention as the independent variable, which could have been specified as a Tier 2, secondary, or targeted intervention or support, provided it was clear the intervention was not a Tier 1 or Tier 3 support. Authors may have explicitly stated the level of intervention. Or, they may have referred to a previous study examining the same intervention in which the intervention was explicitly defined as a Tier 2-level support

Table 1. Coding for Reported Components.

Components	Coding for reported components	Interrater reliability (%)
Treatment integrity (Tier 1)	Reported, not reported, mentioned but not reported, mentioned as a limitation	96.43
Method of measurement	Name and description of methods used to measure treatment integrity	100
Gender	Number of males, females	100
Ethnicity	Number of African American, Asian, Caucasian, Hispanic, Native American, Other	100
Student status	Number of students receiving special education services within specific categories, number of students with mental health diagnosis according to *DSM-IV*, number of students identified at risk of EBD and type of risk based on systematic screening	100
Age	Specific age of participants	92.86
Grade	Specific grade of participants	92.86
Free/reduced lunch or economically disadvantages	Percent of students receiving free/reduced lunch or categorized as economically disadvantaged	96.43
Method of identification	Criteria for participation in study (e.g., systematic screening score, teacher nomination)	89.29
Locale and level	Level (e.g., elementary, middle, high), state or geographic description (e.g., Tennessee, Pacific Northwest), area (e.g., urban, rural, suburban)	100
Treatment integrity (Tier 2)	Reported, not reported, mentioned but not reported, mentioned as a limitation	92.86
Social validity	Reported, not reported, mentioned but not reported, mentioned as a limitation	100
DVs	Name and definition of DV (e.g., problem behavior, correct words per minute)	96.43
Reliability of DVs	Reported, not reported, mentioned but not reported, mentioned as a limitation (e.g., interobserver agreement or reliability statistic)	100
Research design	Description (e.g., single-subject [ABAB, multiple baseline], group [experimental, quasi-experimental, descriptive])	89.29
Intervention domain	Academic, behavioral, and/or social	100
Intervention description	Components (e.g., small group instruction, daily check-in and check-out procedures, differential reinforcement)	100
Reported outcomes	Description based on design (e.g., single-subject designs coded for functional relation, group designs coded for significance of effects)	85.71

Note. DSM-IV = Diagnostic and Statistical Manual of Mental Disorders (4th ed.; American Psychiatric Association, 1994); EBD = emotional/behavioral disorder; DV = dependent variable.

(e.g., Little et al., 2010). If the independent variable was referred to as a manualized or packaged intervention (which are considered Tier 2 supports; Hawken, Vincent, & Schumann, 2008), then the article was included (e.g., Gresham et al., 2006). If the study reported a Tier 2 intervention as part of a larger study of a multileveled prevention model, but the effects of the Tier 2 intervention could not be delineated from overall results, then the study was excluded (e.g., Nelson, Martella, & Marchand-Martella, 2002).

Third, studies employing research designs using single-case or group (i.e., descriptive, quasi-experimental, experimental) methodology were included. For example, single-case studies use (a) the participant as his or her own control, and (b) repeated measurement before, during, and after treatment. Studies were considered descriptive if authors described outcomes of an intervention for a specific participant sample but not in comparison with another participant sample receiving an alternative intervention or no intervention at all. Group designs were considered quasi-experimental if participants were assigned to treatment or comparison groups, but not randomly. In experimental designs, on the other hand, random assignment to treatment and control groups occurred. Nonexperimental studies such as literature reviews, meta-analyses, theoretical or conceptual pieces, qualitative studies, editorials, and case studies were excluded, as they do not provide quantitative evidence regarding intervention outcomes (e.g., Cartledge et al., 2001).

Finally, to ensure the scientific rigor of the studies included for review, only studies published in peer-reviewed

journals were considered. Articles included in non-peer-reviewed journals or books, as well as dissertations, were ineligible for inclusion as the review process for these types of dissemination products vary in rigor.

In sum, of the 67 articles initially retained based on the abstract and then read in entirety for inclusion criteria, 28 studies met the criteria for inclusion in this review. These articles were subsequently coded on 18 components using the following procedures.

Coding Procedures

To systematically identify the presence or absence of various components, as well as to provide descriptive details of components, each study was independently coded by the first and third authors on all 18 components. A final reliability check was done by the second author, resulting in 96.23% agreement (i.e., 18 components × 28 studies = 504 components; agreement on 485 out of 504 components). All discrepancies were resolved to meet 100% agreement. The coded categories included participants, method of identification, setting, intervention components, dependent variable(s), reliability of dependent variable(s), research design, treatment outcomes, treatment integrity, and social validity (see Table 1 for coding description, interrater reliability; see Table 2 for coding of individual studies).

Results

Results were aggregated across the 28 studies to address Tier 1 treatment integrity, participant and setting, measurement, and Tier 2 interventions and their associated outcomes.

Tier 1 Behavioral Plan Treatment Integrity

Tier 1 behavioral plan treatment integrity was reported in 12 of 28 studies. Of these, 10 reported percentages yielded using the Schoolwide Evaluation Tool (SET; Sugai, Lewis-Palmer, Todd, & Horner, 2001). The SET, which has demonstrated internal consistency (Cronbach's $\alpha = .96$) and test–retest reliability at 97.3%, relies on interviews, review of school materials, and observation by an independent assessor to measure treatment integrity (Horner et al., 2004). An overall percentage is calculated for the school (i.e., school as the unit of analysis). Across studies, SET scores ranged from 72% to 100%, with ≥80% suggesting adequate implementation.

Two studies (Lane, Kalberg, Mofield, Wehby, & Parks, 2009; Robertson & Lane, 2007) reported Tier 1 treatment integrity percentages from classroom observation and teacher self-reports. Direct observation and self-report were measured using component checklists of teacher implementation within individual classrooms. Direct observations were completed by research assistants with scores ranging from 74.08% (Lane et al., 2009) to 77.29% (Robertson & Lane, 2007). Teachers' self-report ranged from 78.78% (Lane et al., 2009) to 85.54% (Robertson & Lane, 2007). Direct observation and teacher self-report data were combined to represent a whole-school average of classroom-level implementation.

One study (Wills, Kamps, Abbott, Bannister, & Kaufman, 2010) examining reading interventions within an integrated schoolwide model included a modified version of the Planning and Evaluation Tool (PET) for Effective Schoolwide Reading Programs (Kame'enui & Simmons, 2003). This tool, which contains six items rated as 0, 1, or 2 for a possible 12 points, was used in addition to the SET. School scores ranged from 1 to 12 (Wills et al., 2010).

In addition, findings showed the number of articles measuring and reporting treatment integrity increased over the past 10 years. For example, from 2002 to 2006, there were six studies included in this review with none reporting Tier 1 treatment integrity. From 2007 to 2009, six of 12 (50%) included studies reported Tier 1 treatment integrity. From 2010 to 2011, not only was there an increase in publications ($n = 10$) but also an increase in the percentage of studies reporting Tier 1 plan treatment integrity ($n = 6$, 60%; see Figure 1).

Participants and Setting

Method of identification. Participants in the reviewed studies were deemed eligible for inclusion in their respective studies based on meeting criteria in one or more of the following five categories: office discipline referral (ODR), systematic screening, teacher nomination, academic performance, or other (e.g., parent nomination, attendance, specified behavioral function). Of 28 studies, 10 used ODR data to recruit participants. ODRs were not used in any of these studies as the singular data source for identifying participants. Rather, ODR data were used in conjunction with other indices such as attendance, grades, systematic screeners, or teacher nomination. For example, Hawken and Horner (2003) included participants who had at least five ODRs, were nominated by their teachers, and were not already receiving behavioral support beyond the Tier 1 plan.

Systematic screening was used to identify participants in 13 studies. Many studies using systematic screening data also utilized multiple indices for inclusion. For example, in a study of reading interventions, Wills and colleagues (2011) used Dynamic Indicators for Basic Early Literacy Skills (DIBELS; Kaminski & Good, 1998) and the Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1992) to identify participants with reading and/or behavioral risk factors. Five of these studies used two different screening tools (Gresham et al., 2006; Lane et al., 2008; Lane et al., 2010; Lane et al., 2011; Little et al., 2010).

Table 2. Included Studies.

Components	Lane et al. (2002)	March and Horner (2002)[a]	Hawken and Horner (2003)	Lane et al. (2003)
Treatment integrity (Tier 1)	Not reported	Not reported	Not reported	Not reported
Method of measurement	NA	NA	NA	NA
Gender	Male = 4, female = 3	Male = 20, female = 4	Male = 4	Male = 5, female = 2
Ethnicity	African American = 2, Caucasian = 4, Hispanic = 1	Not reported	20% minority (school)	African American = 2, Hispanic = 5
Student status	SPED = 0 SRSS: risk: moderate = 1, high = 6	SPED = not reported	SPED = 2 (disability not specified)	SPED = 0 SRSS: high risk = 7
Age	6.15 to 8.63	Not reported	12 to 13	8.05 to 9.74
Grade	1st	6th to 8th	6th	2nd to 4th
Free/reduced-price lunch	Not reported	Not reported	31% (school)	Not reported
Method of identification	SRSS, literacy skills	ODR, teacher or parent nomination, 6 weeks attendance	ODR, teacher nomination, not receiving other behavior support	SRSS
Locale and level	Not reported, ES	Not reported, MS	Rural, Pacific NW, MS	Not reported, ES
Treatment integrity (Tier 2)	Reported	Reported	Reported	Reported
Social validity	Reported	Not reported	Reported	Reported
Dependent variable(s)	TDB, NSI, NWF, ORF	Discipline contacts (ODR or detention)	Problem behavior, AE	TDB, NSI, AET
Reliability of DV	Reported	Not reported	Reported	Reported
Research design	Multiple baseline	Descriptive	Multiple baseline	Multiple baseline
Intervention domain	Academic	Behavioral/social	Behavioral/social	Social
Intervention description	Literacy training	BEP	BEP	Social skills instruction
Reported outcomes	Five students completed. TDB decreased for all, NSI decreased for all but one, NWF improved for all, mean ORF improved for all students who completed	12 of 24 participants had lower rates of discipline contacts post-intervention, with the most effective outcomes for behaviors maintained by adult and peer attention	Mean level decreases in problem behavior and increases in AET in all four participants. Researchers reported a "modest" functional relation.	TDB decreased for all groups, NSI decreased for all (except for one participant who had no NSI at baseline), AET increased for all but one participant. Authors reported a functional relation was established for all groups.

Components	Gresham, Van, and Cook (2006)	Hawken (2006)	Fairbanks, Sugai, Guardino, and Lathrop (2007)[b]	Filter et al. (2007)
Treatment integrity (Tier 1)	Not reported	Not reported	Reported	Not reported
Method of measurement	NA	NA	SET	NA
Gender	Male = 2, female = 2	Male = 9, female = 1	Male = 5, female = 5	Not reported
Ethnicity	Not reported	Minority background = 2	African American = 1, Caucasian = 8, Native American = 1	15% to 25% non-White (range across three schools)
Student status	SPED = 0 Mental health: ADHD = 1 (taking 18 mg Concerta) Adapted SSBD: at risk for EBD = 4	SPED = not reported	LD = 2 Mental health: *n* = 2 (taking medication: Concerta, Adderall)	SPED = not reported
Age	6 to 8	Not reported	7 to 8	Not reported
Grade	Not reported	6th to 8th	2nd	Kindergarten to 5th
Free/reduced lunch	Not reported	School = 31%	Not reported	29% to 50% (range across three schools)
Method of identification	Adapted SSBD, SSRS	Entered BEP after 2 months of school, ODR, staff nomination, on BEP for 6 weeks	Teacher nomination	ODR, teacher nomination, no behavior support 6 weeks prior to study
Locale and level	Suburban, S. California, ES	Rural, Pacific NW, MS	Suburban, small city, NW, ES	Pacific NW, ES
Treatment integrity (Tier 2)	Reported	Reported	Reported	Reported
Social validity	Reported	Not reported	Reported	Reported
Dependent variable(s)	TDB, NSI, AT, SSRS	ODR	Problem behavior and AET, teacher perception, ODR	ODR (total, majors, minors)
Reliability of DV	Reported	Not reported	Reported (direct observation data only)	Reported

(continued)

Table 2. (continued)

Components	Gresham, Van, and Cook (2006)	Hawken (2006)	Fairbanks, Sugai, Guardino, and Lathrop (2007)[b]	Filter et al. (2007)
Research design	ABAB withdrawal	Quasi-experimental	Descriptive, quasi-experimental	Quasi-experimental
Intervention domain	Behavioral/social	Behavioral/social	Behavioral/social	Behavioral/social
Intervention description	Social skills instruction, differential reinforcement	BEP	CICO	CICO
Reported outcomes	TDB: functional relation in 3 of 4 NSI: functional relation in 2 of 4 AT: functional relation in 3 of 4 SSRS: means on social skills and problem behavior improved greatly	7 of 10 participants showed reduced ODRs from baseline to implementation of BEP	4 of 10 had decreased problem behavior, improved teacher perception of behavior, and fewer ODR per day indicating responsiveness to intervention. Functional relation not established due to the design.	Significant decrease in total ODR for 12 students; decrease in major offenses, but not significant; significant decrease in minor offenses

Components	Hawken, MacLeod, and Rawlings (2007)	M. R. Marchant et al. (2007)	Robertson and Lane (2007)	Campbell and Anderson (2008)
Treatment integrity (Tier 1)	Reported	Not reported	Reported	Not reported
Method of measurement	SET	NA	Direct observation, teacher self-report	NA
Gender	Male = 10, female = 2	Male = 2, female = 1	Male = 42, female = 23	Male = 2
Ethnicity	Minority: 38% (school), *n* = 2	Caucasian = 3	Caucasian = 63, Hispanic = 2	African American = 1, Caucasian = 1
Student status	LD = 1	SPED = 0 SSBD: internalizing = 3	SPED = 17 (categories not reported) SRSS: Moderate/high risk (*n* = not reported)	SPED = 0
Age	Not reported	7, 11, 11	11.48–15.48	10
Grade	Not reported	1st and 5th	6th to 8th	Not reported
Free/reduced lunch	66% (school), 67% (participants)	50% (school), lower to middle class SES (participants)	29% (school)	Not reported
Method of identification	ODR, on BEP after 2 months of school, on BEP for 6 weeks, teacher nomination, and display problem behavior	SSBD, informal observations	SRSS, ODR, GPA, course failures	Teacher or team nomination
Locale and level	Urban, ES	Central Utah, ES	Rural, Middle TN, MS	Rural, Pacific NW, ES
Treatment integrity (Tier 2)	Reported	Reported	Reported	Reported
Social validity	Reported	Reported	Reported	Not reported
Dependent variable(s)	ODR (total, majors, minors)	Effective communication, appropriate peer-play	KSS, KCRS, SHI, CT	Problem behavior
Reliability of DV	MNR	Reported	Reported	Reported
Research design	Multiple baseline	Multiple baseline	Experimental	ABCBC reversal
Intervention domain	Behavioral/social	Behavioral/social	Academic/social	Behavioral/social
Intervention description	BEP	Social skills training, peer/adult mediation, self-management system with positive reinforcement	Study skills instruction (SS), or study skills and conflict resolution instruction (SS + CR)	CICO with function-based adaptation
Reported outcomes	Mean level decreases in total ODR per month across all groups, demonstrating a functional relation 75% decrease in average referrals per month. Ten students required additional support.	Functional relation established between treatment package with peer mediation and DVs, but adult mediation was even more effective than peer mediation	Both groups (SS and SS + CR) increased on KSS, but did not achieve mastery. SS + CR group increased on KCRS, but did not achieve mastery. Both groups showed higher postintervention scores on SHI and CT.	CICO with function-based support produced greater changes than CICO-alone.

Components	Cheney, Flower, and Templeton (2008)	Lane et al. (2008)	Todd, Campbell, Meyer, and Horner (2008)	Cheney et al. (2009)
Treatment integrity (Tier 1)	Not reported	Not reported	Reported	Not reported
Method of measurement	NA	NA	SET	NA
Gender	Male = 100, female = 27	Male = 5, female = 1	Male = 4	Male = 152, female = 55
Ethnicity	African American = 25, Asian = 8, Caucasian = 78, Hispanic = 12, Native American = 4	African American = 2, Caucasian = 4	African American = 1, Caucasian = 2, Native American = 1	African American = 36, Asian = 16, Caucasian = 107, Hispanic = 23, Native American = 1

(continued)

Table 2. (continued)

Components	Cheney, Flower, and Templeton (2008)	Lane et al. (2008)	Todd, Campbell, Meyer, and Horner (2008)	Cheney et al. (2009)
Student status	SPED = 23 (not reported = 9) OHI = 2, DD = 6, SLI = 1, ASD = 1, LD = 3, EBD = 1 SSBD: externalizing = 94, internalizing = 33	SPED = not reported SSBD: externalizing = 2, internalizing = 1 SRSS Risk: moderate = 2, high = 3	SPED = 1 (category not reported) Mental health: *n* = 1 (taking 20 mg Ritalin)	SPED = 48 (categories not reported) SSBD: Ext/internalizing = 207 (type not reported)
Age	Not reported	7.5 to 8.3	Not reported	Not reported
Grade	1st to 3rd	2nd	K to 3rd	1st to 5th
Free/reduced lunch	11.6% to 77% (schools)	Not reported	Not reported	*n* = 48, 53% (district percentage)
Method of identification	SSBD	SSBD, SRSS, TOWL-3	Administrator nomination, teacher verification	SSBD, IEP
Locale and level	Puget sound and SW regions, Washington, ES	Rural, Middle TN, ES	Rural, Pacific NW, ES	Western Washington, ES
Treatment integrity (Tier 2)	Reported	Reported	ML	Reported
Social validity	Not reported	Reported	Reported	Reported
Dependent variable(s)	DPR, SSRS	Writing probes (elements, words, quality)	Problem behavior, ODR	SSRS, TRF, AET, WJIII
Reliability of DV	Reported	Reported	Reported (problem behavior), not reported (ODR)	Reported
Research design	Design not stated	Multiple-baseline/multiple-probe	Multiple baseline	Experimental
Intervention domain	Behavioral/social	Academic/behavioral	Behavioral/social	Behavioral/social
Intervention description	CCE	SRSD, reinforcement	CICO	CCE
Reported outcomes	Responders: (a) absolute change: 88 of 127, (b) PND: inadequate measurement, (c) % change: 76 of 127, (d) ES in positive direction for 88 of 127, (e) RCI: 50 of 127, and (f) 38 of 68 "true responders" made positive gains on SSRS behavior scale.	All students were highly responsive: 100% PND on story elements, large increases in number of words, improved quality. Functional relation established. Students maintained gains during follow-up.	Functional relation between intervention and decreases in problem behavior for three of four participants.	60% graduated and moved into normal range of problem behavior on SSRS and TRF. Graduates significantly decreased problem behavior (SSRS) and showed significantly lower externalizing and internalizing behavior scores (TRF) in comparison to control and nongraduate groups. No statistically reliable effects on AET or WJIII.

Components	Lane, Kalberg, Mofield, Wehby, and Parks (2009)	McIntosh, Campbell, Carter, and Dickey (2009)	Lane et al. (2010)	Little et al. (2010)
Treatment integrity (Tier 1)	Reported	Reported	Not reported	Not reported
Method of measurement	Direct observation, teacher self-report	SET	NA	NA
Gender	Male = 66, female = 60	Male = 28, female = 6	Male = 8; female = 5	Male = 7, female = 6
Ethnicity	African American = 4, Caucasian = 120, Hispanic = 2	African American = 3%, Caucasian = 88%, Hispanic = 6%, Native American = 3%	Caucasian = 13	Caucasian = 13
Student status	LD = 9, Gifted = 8, OHI = 1	SPED = not reported	SPED = 1 (category not reported) SSBD: externalizing = 4, internalizing = 6; SRSS Risk: moderate = 5, high = 5	OHI = 1, OI = 1 SSBD: externalizing = 7, internalizing = 6, SRSS risk = moderate = 4, high = 8
Age	Not reported	6 to 11	7.03 to 9.01	7.01 to 8.09
Grade	11th	1st to 5th	2nd	2nd
Free/reduced-price lunch	Not reported	53% (district)	4.8% to 19.5% (schools)	4.8% to 19.5% (schools)
Method of identification	Principal and schoolwide team	Teacher nomination	SRSS, SSBD, TOWL-3	SRSS, SSBD, TOWL-3
Locale and level	Rural, Middle TN, HS	Pacific NW, ES	Rural, Central TN, ES	Rural, Mid-South, ES
Treatment integrity (Tier 2)	Reported	Reported	Reported	Reported
Social validity	Reported	Not reported	Reported	Reported
Dependent variable(s)	ACT actual test, ACT practice test	BASC-2, ODR	Writing probes (elements, words, quality)	Writing probes (elements, words, quality)
Reliability of DV	Reported	Reported	Reported	Reported
Research design	Descriptive, quasi-experimental	Quasi-experimental	Multiple-baseline/multiple-probe	Multiple-baseline/multiple-probe
Intervention domain	Academic	Behavioral/social	Academic/behavioral	Academic/behavioral

(continued)

Table 2. (continued)

Components	Lane, Kalberg, Mofield, Wehby, and Parks (2009)	McIntosh, Campbell, Carter, and Dickey (2009)	Lane et al. (2010)	Little et al. (2010)
Intervention description	ACT preparation curriculum	CICO	SRSD, reinforcement	SRSD, reinforcement
Reported outcomes	Only GPA in previous academic year predicted postintervention practice scores. Postintervention practice ACT scores predicted actual ACT scores. GPA and ODR data a predicted actual English ACT scores. Participants meeting the district target increased 10% from previous year. They met or exceeded state means in all subtests except for reading.	Participants with attention function decreased problem behavior ratings, increased prosocial ratings, and decreased ODR. Participants with task avoidance function had no significant changes in problem behavior or prosocial ratings, and no decrease in ODR.	For students with externalizing behavior: all improved in writing elements, 6 in total words, and 6 in quality. For students with internalizing behavior: all improved in elements and total words, with 4 in quality. Functional relation between number of elements and SRSD for 7 students with externalizing and 5 with internalizing behaviors.	Students with externalizing behaviors improved in writing elements, quality, and total words. All students with internalizing behavior improved in elements and five students improved in words and quality. Functional relation between number of elements and SRSD for all students at risk of writing difficulties.

Components	Wills, Kamps, Abbott, Bannister, and Kauffman (2010)	Campbell and Anderson (2011)	Hawken, O'Neil, and McLeod (2011)	Kamps et al. (2011)
Treatment integrity (Tier 1)	Reported	Reported	Reported	Reported
Method of measurement	SET, PET	SET	SET	SET
Gender	Male = 101, female = 70	Male = 4	Male = 11, female = 6	School 1 = not reported, School 2: Male = 10, female = 6, School 3 = not reported, at risk of EBD: Male = 5, female = 3
Ethnicity	African American = 38%, Caucasian = 32%, Hispanic = 23%	Caucasian = 3, Hispanic = 1	Caucasian = 14, ethnic minorities = 3	Minority: 18% to 90% (schools)
Student status	EBD = 1, LD = 24, SLI = 4 SSBD: externalizing = 87	LD = 2, SLI = 1 Mental health: ADHD = 2 (taking methyphenidate, amphetamine, desxtroamphetamine)	SPED = not reported	SPED = not reported SSBD: At risk of EBD = 8
Age	Not reported	Not reported	Not reported	Not reported for whole sample, At risk of EBD = 9 to 11
Grade	1st to 3rd	2nd and 5th	1st, 2nd, 3rd, 5th, 6th	K, 1st, 4th, 5th
Free/reduced-price lunch	School ranges (*n* = 8): 13% to 97%	School = 72%	School 1 = 63%, School 2 = 47%	20% to 94% (schools)
Method of identification	DIBELS, SSBD	ODR, school team, teacher nomination	On BEP after first 2 weeks of school, ODR, teacher nomination, received BEP intervention for at least 4 weeks, demonstrated problems in >1 academic period	Classroom = voluntary, Students = teacher nomination, SSBD
Locale and level	Urban, suburban, Midwestern metropolitan area, ES	Pacific NW, ES	Urban, ES	School 1: urban ES, School 2: suburban ES, School 3: low-income community, ES
Treatment integrity (Tier 2)	Reported	Reported	Reported	Reported
Social validity	Not reported	Reported	Reported	Reported
Dependent variable(s)	NWF, ORF, WRMT, Reading time, academic responding (compliance), student conduct, instructional grouping minutes, teacher praise and reprimands	Problem behavior, AET, DPR	ODRs, referrals for SPED or additional behavior support	On-task, disruptive behavior, teacher praise, and reprimands
Reliability of DV	Reported	Problem behavior = reported; AET = reported; percentage of points earned = not reported	ODR: MNR, SPED: not reported	Reported
Research design	Quasi-experimental	Reversal plus component analysis	Quasi-experimental	Reversal with variations across classrooms
Intervention domain	Academic	Behavioral/social	Behavioral/social	Behavioral/social

(continued)

Table 2. (continued)

Components	Wills, Kamps, Abbott, Bannister, and Kauffman (2010)	Campbell and Anderson (2011)	Hawken, O'Neil, and McLeod (2011)	Kamps et al. (2011)
Intervention description	Multiple-reading interventions	CICO	BEP	CW-FIT
Reported outcomes	Students in experimental groups averaged more time reading in small groups, reading aloud, and had twice the responding rates of students in control groups. Control and experimental groups had similar rates of disruption, with students exhibiting behavior/reading risk displaying higher rates than those with only reading risk. Experimental group had significantly higher growth rates on DIBELS for first grade (NWF, ORF) and third grade (ORF). Experimental group did significantly better on Word ID subtest of WRMT. No significant differences on Word Attack or Passage Comprehension Tests on WRMT.	Increases in AET and decreases in problem behavior. Functional relation established for all participants. Slight reduction in percentage of points earned as feedback sessions were removed.	Significant decrease in ODRs across schools, but effectiveness varied slightly by function and severity of behavior. One student referred for SPED and qualified under OHI. One student referred for more behavioral support.	Increased on-task behavior (ES = 2.03) for all classes and grades, including students at risk of EBD. For at-risk students, disruptive behavior decreased (ES = 1.29), on-task behavior increased (ES = 0.93). Functional relation between CW-FIT and (a) on-task behavior in four of six classes, and (b) praise and reprimands in three of five classes.

Components	Lane et al. (2011)	Mong, Johnson, and Mong (2011)	Ness, Sohlberg, and Albin (2011)	Simonsen, Myers, and Briere (2011)
Treatment integrity (Tier 1)	Not reported	Reported	Not reported	Reported
Method of measurement	NA	SET	NA	SET
Gender	Male = 32, female =12	Male = 2, female = 2	Male = 3	Male = 32, female = 10
Ethnicity	African American = 1, Caucasian = 41, Hispanic = 1	African American = 2, Caucasian = 2	Caucasian = 80% (district)	African American = 3, Caucasian = 5, Hispanic = 33, Native American = 1
Student status	ASD = 1, LD = 2, OHI = 4, SLI = 1, SSBD: Ext = 24, Int = 20, SRSS: Low = 16, Mod = 15, High = 13	SPED = 0	ASD = 1, EBD = 1, LD = 1	SPED = 7 (categories not reported)
Age	6.71 to 8.50	8.4 to 9.1	Not reported	10 to 14
Grade	2nd	3rd	7th	5th to 8th
Free/reduced-price lunch	6.9% to 34.5% (schools)	65% (school)	45% (district)	71% (school)
Method of identification	SRSS, SSBD, TOWL-3	ODR, behavior across multiple settings, function was attention	Receiving resource room support, low assignment attack behavior	ODR, teacher nomination
Locale and level	Rural, Southeast, ES	Suburban, Southeast, ES	Urban, Pacific NW, MS	Urban, New England, MS
Treatment integrity (Tier 2)	Reported	Reported	Reported	Reported
Social validity	Reported	Reported	Reported	Reported
Dependent variable(s)	Writing probes (OE & SW, teacher, and RA administered), AET, SSRS-T, SSRS-P	Problem behavior, ODR, DPR, DCPM, EPM	Assignment attack behavior	Off-task, SSRS
Reliability of DV	Reported	Problem behavior = reported, ODR: MNR, DPR: MNR, DCPM: reported, EPM: Reported	Reported	Off-task = reported, SSRS = reported
Research design	Experimental	Combined series multiple baseline	ABAB reversal	Experimental
Intervention domain	Academic/behavioral	Behavioral/social	Behavioral	Behavioral/social
Intervention description	SRSD, reinforcement	CICO	Assignment management package: organization and self-monitoring	BEP

(continued)

Table 2. (continued)

Components	Lane et al. (2011)	Mong, Johnson, and Mong (2011)	Ness, Sohlberg, and Albin (2011)	Simonsen, Myers, and Briere (2011)
Reported outcomes	OE: Tx did significantly better than control for elements, quality, & length on RA probes (ES = 1.08 to 1.66), but no significant differences on teacher-administered probes. SW: Tx did significantly better than control on RA (ES = 1.12) and teacher (ES = 1.04) probes for elements and RA probes for quality (ES = 1.20), with no other significant improvements on SW. AET: Tx had significant increases during OE (ES = .84), no significant increases during SW. No significant differences on SSRS-T or SSRS-P.	Decreased problem behavior and ODR, increased DPR, increased DCPM, and decreased EPM for all participants. Despite positive outcomes, no functional relation established due to experimental design.	Functional relation established for two of three participants.	BEP resulted in significantly less off-task behavior than standard practice (SP; [control group]), ES = −.90, No significant differences on SSRS gain scores (i.e., problem behavior [ES = .65], social skills [ES = −.12], and academic competence [ES = −.17]) between groups, but ES's favored SP condition.

Note. NA = not applicable; SPED = receiving special education services; SRSS = Student Risk Screening Scale; ODR = office discipline referral; ES = elementary school; MS = middle school; TDB = total disruptive behavior; NSI = negative social interactions; AET = academic engaged time; NWF = nonsense word fluency; ORF = oral reading fluency; DV = dependent variable; BEP = behavior education program; AE(T) = academic engagement (time); SSBD = Systematic Screening for Behavior Disorders; EBD = emotional/behavioral disorder; AT = alone time; SSRS = social skills rating system; SET = schoolwide evaluation tool; LD = learning disability; CICO = check-in/check-out; GPA = grade point average; KSS = knowledge of social skills; KCRS = knowledge of conflict resolution skills; SHI = study habits inventory; CT = conflict talk; OHI = other health impairment; DD = developmental delay; SLI = speech/language impairment; ASD = Autism spectrum disorder; TOWL-3 = Test of Oral and Written Language-3; ML = mentioned as a limitation; DPR = daily progress report; CCE = check, connect, and expect; SRSD = self-regulated strategy development; PND = percentage of non-overlapping data; RA = research assistant; RCI = reliable change index; IEP = Individualized Education Program; HS = high school; TRF = teacher report form; Tx = treatment group; WJIII = Woodcock–Johnson III Tests of Achievement; ACT = Academic College Testing; OI = orthopedic impairment; BASC-2 = Behavior Assessment Scale for Children-2; DIBELS = Dynamic Indicators for Basic Early Literacy Skills; WRMT = Woodcock Reading Mastery Test; MNR = mentioned but not reported; DCPM = digits correct per minute; EPM = Errors per minute; CW-FIT = classwide function-based intervention teams; OE = opinion essay; SW = story writing; O = Other.
[a]Only results of descriptive analyses are reported as the function-based interventions were considered Tier 3 support. [b]Only results of Study 1 are reported, as Study 2 reflects Tier 3 support.

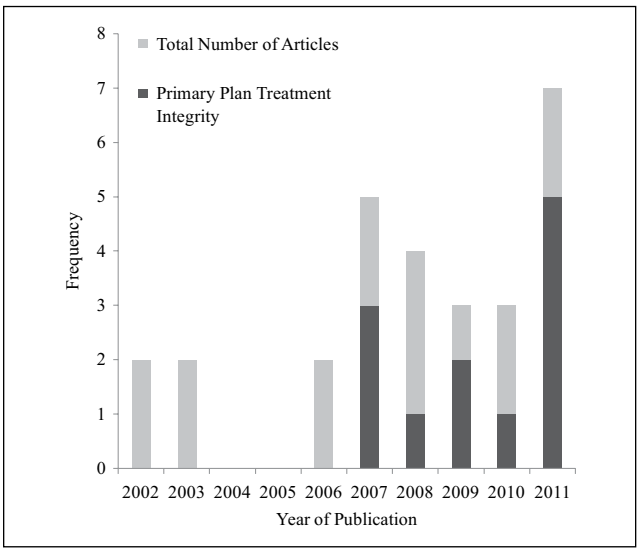

Figure 1. Primary plan treatment integrity by year of publication.

In 11 studies, participants were identified by teacher nomination alone or in conjunction with other criteria. Descriptions of teacher nomination procedures varied and did not include a quantifiable index. For example, McIntosh, Campbell, Carter, and Dickey (2009) reported that students were nominated by classroom teachers based on their perception of students' need for additional support beyond the primary prevention plan.

In seven studies, participants were identified by academic data such as grade point average or course failure (e.g., Robertson & Lane, 2007), Test of Oral and Written Language-3 (TOWL-3) scores (Hammill & Larsen, 1996; e.g., Little et al., 2010), or DIBELS scores (Kaminski & Good, 1998; e.g., Wills et al., 2010). All studies using an academic measure also included a behavioral measure (e.g., ODR).

Demographic characteristics. Gender was reported in all but 1 study (Filter et al., 2007). In the 27 studies reporting gender, over twice as many participants were male (n = 675) compared with 312 females.

Ethnicity was reported nearly as frequently as gender (n = 26). In 5 studies, ethnicity was reported as either a whole school or district percentage. In the 19 studies reporting ethnicity of individual participants, 481 were Caucasian, 80 African American, 80 Hispanic, 24 Asian/Pacific Islander, and 8 Native American.

In 19 studies, school or district percentage of students being economically disadvantaged or receiving free or reduced-price lunch was reported. In addition, one study

reported an individual as being economically disadvantaged and another study reported an average across participants.

All but three studies reported participant grade level (Campbell & Anderson, 2008; Gresham et al., 2006; Hawken, MacLeod, & Rawlings, 2007). Grade levels ranged from kindergarten to 11th grade. Age or age ranges were explicitly stated in more than half of the articles (n = 15). Reported ages ranged from as young as 6 years to as old as 15 years.

Seven studies did not include disability status of participants. It is not known whether these data were not reported because none of the participants had a disability, if the data were not available to researchers, or if they simply were not included in the study. In the 21 studies reporting disability status, 44 students had specific learning disability, 7 had speech/language impairments, 3 had autism spectrum disorder, 3 had emotional disturbance, 8 were identified as gifted, 1 had an orthopedic impairment, 8 received services under the "other health impairment" (OHI) category, and categories were not specified for 86 participants.

In 12 studies, students were identified as at risk of emotional or behavioral disorders (EBD) by a systematic screener. Specifically, using the Student Risk Screening Scale (SRSS; Drummond, 1994), 27 participants were found at moderate risk (total score = 4–8) and 42 were at high risk (total score = 9–21) for antisocial behavior. Using the SSBD, 218 participants had critical externalizing behaviors and 69 had internalizing behaviors. For 219 participants, the SSBD was used to identify risk for EBD, but the type (i.e., internalizing or externalizing) was not specified.

Only four studies included participants with mental health concerns according to *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM-IV-TR*; American Psychiatric Association, 2000) criteria (Campbell & Anderson, 2011; Fairbanks, Sugai, Guardino, & Lathrop, 2007; Gresham et al., 2006; Todd, Campbell, Meyer, & Horner, 2008). Of these, three participants had attention deficit-hyperactivity disorder (ADHD) and six participants were taking medication (e.g., ritalin, concerta, adderall).

Locale and level. Locale, which was reported in 25 of 28 studies, varied from rural (n = 10) to suburban (n = 4) to urban (n = 6). In addition, 19 studies took place in either the Pacific West (e.g., Washington and California; n = 12) or the South (e.g., Tennessee; n = 7). The majority of studies took place at the elementary level (n = 21), with 6 at the middle school level, and only 1 at the high school level.

Measurement

Treatment integrity. Tier 2 intervention integrity was frequently reported, with all but 1 study including Tier 2 treatment integrity data (Todd et al., 2008). These authors described the lack of treatment integrity data as a limitation of the study (Todd et al., 2008). Across 27 studies, treatment integrity was measured using permanent products from the intervention (e.g., Hawken et al., 2007) or procedural checklists completed by interventionists (e.g., Filter et al., 2007) and outside observers (e.g., Cheney, Flower, & Templeton, 2008). In these studies, procedural components of interventions were implemented with moderate to high levels of fidelity.

Social validity. Social validity was measured and reported frequently (n = 22). Only six studies did not include information related to the acceptability of the goals, procedures, and outcomes associated with intervention (Campbell & Anderson, 2008; Cheney et al., 2008; Hawken, 2006; March & Horner, 2002; McIntosh et al., 2009; Wills et al., 2010). Social validity was assessed with student, parent, and teacher questionnaires (e.g., Hawken & Horner, 2003); interviews (e.g., Lane et al., 2002); rating scales (e.g., Lane et al., 2009); and measures of behavior change (e.g., Gresham et al., 2006). In the studies reporting social validity data, interventions were judged as mostly favorable. Specifically, teachers and/or students rated the intervention favorably (e.g., Cheney et al., 2009), would recommend it to others (e.g., Hawken & Horner, 2003), found procedures acceptable (e.g., Todd et al., 2008), and reported socially important, measurable changes in student behavior (e.g., Gresham et al., 2006).

Dependent variable(s). Dependent variables (DVs) included academic measures (e.g., oral reading fluency, Academic College Testing [ACT] score, number of writing elements) and behavioral measures utilizing direct observation, rating scales, and teacher reports. For interventions targeting academic skills, most DVs were proximal measures of intervention effects. For example, in a study targeting literacy skills, students were assessed on nonsense word fluency and oral reading fluency (Lane et al., 2002). The most commonly used DVs used to measure behavior change were ODRs (n = 9), problem behavior (n = 6), and on-task/academic engagement (n = 7). Generally, ODRs were used in group design studies, whereas direct observations of behavior (e.g., disruptive behavior, negative social interactions [NSI]) were used as DVs in single-subject design studies.

To be sure changes in academic, behavioral, and social skills were due to intervention and not due to measurement error, we recorded the presence of dependent variable reliability. Across all studies, DV reliability was reported 63 out of 74 possible times, with four studies mentioning but not reporting reliability and eight studies neither mentioning nor reporting reliability. For example, NSIs were used 3 times as a DV, and each time the reliability was reported. On the other hand, ODRs were used as the DV in nine different studies and in two studies reliability was reported, in three studies reliability was mentioned but not reported, and

in four studies there was no mention or reporting of reliability. In studies mentioning but not reporting reliability estimates for ODRs, authors described ODRs as reliable and valid based on previous research (e.g., Hawken, O'Neil, & McLeod, 2011; Mong, Johnson, & Mong, 2011). In all studies involving direct observation of behavior, reliability was reported as a percentage of interobserver agreement (IOA) with (a) methods for calculating percentages varying based on the type of observational recording system (e.g., interval, duration, frequency, rating scale), and (b) percentages reported either by participant, dependent variable, or as an overall average across participants or dependent variables. Only one study included a kappa value (.89; Hawken & Horner, 2003).

Research design. Single-subject designs, in which behavior was directly observed, were the most frequently used experimental method (n = 16). Most single-subject studies utilized multiple-baseline designs (n = 7), with four studies using a multiple-baseline/multiple-probe design, and five using a reversal or withdrawal design. All single-subject studies included a graph for visual analysis to determine the presence of a functional relation.

In terms of group design methodology, authors described studies as being descriptive, quasi-experimental, or experimental in nature. Of the 14 group design studies, 3 were descriptive, 7 were quasi-experimental, and 4 were experimental. The most rigorous evaluations were employed in studies using experimental, pre–post designs with random assignment to intervention and control conditions. For example, stratified randomization was used in an investigation by Cheney and colleagues (2009) in which schools were matched on four variables and then randomly assigned to either intervention or comparison conditions.

Tier 2 Interventions and Outcomes

Interventions targeted either singular academic, behavioral, or social skills domain, or a combination of these domains. Intervention outcomes were mostly positive.

Behavioral/social. Almost two thirds of studies targeted behavioral and social skills (n = 18). In 15 of 18 studies, these skills were addressed using Check, Connect, and Expect (CCE), CICO, or the Behavior Education Program (BEP), which share the common component of students meeting with an adult mentor at the beginning of the day to set behavioral goals and end of the day to receive feedback about those goals. In the majority of these studies, students demonstrated decreases in ODR rates (e.g., McIntosh et al., 2009), increases in academic engagement (e.g., Hawken & Horner, 2003), and decreases in problem behaviors (e.g., disruption, NSI; e.g., Campbell & Anderson, 2008). Furthermore, visual analysis of graphs revealed functional relations between CICO variations (i.e., BEP, CCE) and observed behaviors for most, but not all, participants in 5 of 7 studies using single-case design (Campbell & Anderson, 2008, 2011; Hawken et al., 2007; Hawken & Horner, 2003; Todd et al., 2008). In 2 studies, a functional relation could not be established due to design limitations (Fairbanks et al., 2007; Mong et al., 2011). In descriptive investigations of CCE, CICO, and BEP, significant differences were noted in participants' daily progress reports (Cheney et al., 2008), standardized behavioral ratings (Cheney et al., 2009; McIntosh et al., 2009), and ODRs (Filter et al., 2007; March & Horner, 2002; McIntosh et al., 2009). However, not all participants were responsive to CCE, CICO, or BEP. For example, in a descriptive study of CICO, McIntosh and colleagues (2009) found significant differences in BASC-2 scores and ODRs for only participants with attention-motivated behavior. Students whose behavior served to avoid tasks showed no significant changes in BASC-2 ratings or ODRs. Likewise, Hawken and colleagues (2011) found the effectiveness of BEP varied by function as well as severity of behavior. March and Horner (2002) investigated the effects of BEP and found only half of the participants (i.e., 12 of 24) had lower rates of discipline contacts.

In addition to CCE, CICO, and BEP studies, three other studies reported outcomes of a social-behavioral intervention. Using a single-subject reversal design, Kamps et al. (2011) reported a functional relation between classwide function-based intervention teams (CW-FIT) and on-task behavior in four of six classes, as well as praise and reprimands in three of five classes. Students identified as at risk for EBD also demonstrated significant decreases in disruptive behavior (elementary school [ES] = 1.29) and increases in on-task behavior (ES = 0.93). Gresham et al. (2006) used a withdrawal design to determine the effects of social skills instruction in conjunction with differential reinforcement of other behavior. Instruction occurred 2 days per week for 20 weeks with sessions lasting 1.5 hr each. Results revealed functional relations between intervention and (a) disruptive behavior in three of four participants, (b) NSI in two of four participants, and (c) alone time in three of four participants. In addition, all four participants improved scores in social skills, problem behavior, and academic competence on the Social Skills Rating System (SSRS; Gresham & Elliott, 1990). Similarly, M. R. Marchant et al. (2007) reported a functional relation between a social skills intervention package (i.e., social skills training, peer and adult mediation, self-management, reinforcement) and the communication and play skills of three students with internalizing behaviors.

Social. Only one study focused on the social domain alone. In the study by Lane et al. (2003), explicit instruction in social skills was delivered to students in 30-min sessions, 2 days per week for 10 weeks. Using a multiple-baseline design, authors reported a functional relation between social

skills instruction and DVs (i.e., total disruptive behavior [TDB], NSI, academic engaged time [AET]) for all groups (Lane et al., 2003).

Social/academic. Only one study occurred in the combined social and academic domains (Robertson & Lane, 2007). In this study, participants were randomly assigned to manualized study skills instruction or study skills plus conflict resolution instruction with interventions occurring in group settings. Students receiving study skills instruction alone and study skills plus conflict resolution instruction had low magnitude increases in their knowledge of study skills, but they did not achieve mastery of these skills. Findings indicated that participants' study habits did not significantly improve for either group. The study skills plus conflict resolution improved their knowledge of conflict resolution skills, but they did not achieve mastery. Again, this knowledge did not result in improved conflict resolution styles.

Academic. Three studies targeted only academic skills. For example, using a multiple-baseline design across interventions groups, Lane and colleagues (2002) found literacy instruction contributed to decreases in TDB for all five participants and decreases in NSI for all but one participant. Moreover, nonsense word fluency (NWF) improved for all participants and oral reading fluency (ORF) improved for three of five. Data were graphed for only two of the five participants, with a functional relation established between literacy training and reading (NWF) and behavioral (TDB) outcomes for both illustrations.

Similarly, in a quasi-experimental study examining the effects of different manualized Tier 2 reading interventions, students in intervention had significantly higher growth rates on DIBELS (Wills et al., 2010). For first graders, higher growth rates were demonstrated in NWF and ORF. For third graders, higher growth rates were demonstrated in ORF. However, there were no significant differences between experimental and control groups on Word Attack or Passage Comprehension subtests of the *Woodcock Reading Mastery Test*.

The third study in the academic domain examined the effects of an ACT preparation curriculum and found participants meeting the district target for ACT score increased 10% from the previous year (Lane et al., 2009). As a group, participants also met or exceeded state means in all ACT subtests except reading, which was an improvement from the previous year's class who did not receive the curriculum and scored below the state mean on all ACT subtests and total score.

Academic/behavioral. Three studies employed multiple baseline/multiple-probe designs to examine the effects of self-regulated strategy development (SRSD) for writing plus planned reinforcement on students with critical internalizing or externalizing behaviors (Lane et al., 2008; Lane et al., 2010; Little et al., 2010). Functional relations were established between intervention and the number of story elements, number of words, and writing quality scores for the vast majority of participants. Although the intervention targeted students with writing and behavioral difficulties, only dependent measures of writing were included. However, the most recent study of SRSD and reinforcement did include writing and behavioral dependent variables (Lane et al., 2011). Results of this randomized control trial demonstrated the treatment group (i.e., SRSD plus reinforcement) did significantly better than the control group on research assistant-administered opinion essay probes in terms of elements, quality, and length (ES = 1.08 to 1.66), but no statistically significant differences on teacher-administered probes. In terms of story writing probes, results indicated the treatment group significantly better than control on research-administered (ES = 1.12) and teacher-administered probes for elements (ES = 1.04) and research assistant-administered probes for quality (ES = 1.20; see Table 2). The treatment group demonstrated significant increases in academic engagement (ES = .84) during opinion essay probes, but not during story writing probes. Additionally, there were no significant improvements on the problem behavior subscale of the SSRS (parent and teacher versions).

Behavioral. Ness, Sohlberg, and Albin (2011) examined the effects of an assignment management intervention package on the effects of assignment attack behavior of three students with disabilities who lacked organization and task persistence. Using a reversal design, a functional relation between the packaged intervention and assignment attack behavior was established for all three participants.

To summarize, interventions focused on a range and combination of academic, behavioral, and social skills. The largest percentage of studies were conducted within the behavioral-social domain (64.29%), with more than half of all studies examining variations of CICO (i.e., BEP, CCE; $n = 15$), and the majority of studies reporting positive outcomes.

Discussion

In multitiered prevention models, practitioners aim to create a secure and productive learning environment that is responsive to all students' needs. As part of this aim, researchers, teachers, and administrators have collaborated to determine effective ways to intervene with students for whom the Tier 1 plan is insufficient. In this review, we were challenged to determine whether Tier 2 interventions were warranted based on nonresponsiveness to Tier 1, the participants and settings of Tier 2 intervention research, the types of interventions that have been implemented and evaluated, as well as their associated outcomes. In the

remaining text, we summarize key findings and reflect on limitations associated with each finding by offering suggestions for future inquiry.

Tier 1 Behavioral Plan Treatment Integrity: A Foundational Component

Before decisions can be made about allocating time, money, and personnel to Tier 2 interventions, schools should ensure the Tier 1 plan is being implemented as designed. Only then can an accurate assessment of a student's responsiveness to intervention occur and resources be allocated for more intense supports or professional development in Tier 1 implementation. In this review, only 12 of 28 studies reported Tier 1 plan treatment integrity, which is a distinct limitation of the literature base. As the field has begun to recognize the importance of treatment integrity, reporting Tier 1 integrity has certainly increased over the last several years (see Figure 1). Whereas 10 of 12 studies included the SET to measure school-level Tier 1 integrity, 2 studies used direct observation and self-report of Tier 1 implementation within individual classrooms (Lane et al., 2009; Robertson & Lane, 2007). In these 2 studies, however, data were aggregated and reported as averages of all teachers' implementation rather than individual percentages of the participating students' teachers. Although this is a positive stride to better understanding Tier 1 behavioral plan implementation within classrooms, we recommend reporting Tier 1 plan treatment integrity for each participant's teacher to better understand whether a Tier 2 intervention is justified. By measuring Tier 1 plan treatment integrity at the classroom level, schools can (a) identify students who are truly nonresponsive to the Tier 1 plan and (b) ensure limited resources are allocated appropriately for Tier 2 supports.

Because all of these tools (e.g., SET, direct observation checklists, self-report checklists) provide unique and valuable information, we agree with recent recommendations for taking a multi-method, multi-informant approach to measuring Tier 1 treatment integrity (Keller-Margulis, 2012; Roach & Elliott, 2008). Schools could benefit from using more than one measure from more than one perspective. This approach should involve measuring treatment integrity not only of instructional components (e.g., teaching behavioral expectations), which are included in all measures reported in this review, but also data-based decision-making processes as well (Roach & Elliott, 2008). For example, future research may involve measuring the integrity of the data-based procedures used to identify nonresponders and match them with an appropriate intervention.

Students Receiving Tier 2 Interventions

The difference in approaches to identifying students for Tier 2 supports is evidenced in this body of literature with roughly one third of included studies using ODRs and almost half of the studies using systematic screeners. However, in all studies using ODRs or systematic screeners, other data sources also were used as inclusion criteria for Tier 2 intervention. These data often included academic measures or teacher nomination. The key to identification is having measures that paint an accurate picture of the student. Recently, M. Marchant et al. (2009) suggested triangulating ODRs, systematic screeners, and qualitative data (e.g., surveys, faculty discussions) to refine and develop Tier 1 interventions. Similarly, we recognize the benefit of data triangulation in identifying and supporting students for Tier 2 interventions. The idea is that schools can target specific academic deficits through Tier 2 interventions, as well as provide the behavioral support necessary for students to access the academic instruction based on data gleaned from multiple sources (Lane, Menzies, Oakes, & Kalberg, 2012; McIntosh, Sadler, & Brown, 2012). This is particularly important given the limitations of using ODRs or systematic screeners alone. For example, there is still concern about (a) the variation among teachers in total ODRs generated, (b) the inability of ODRs to identify students with internalizing behaviors, and (c) the "wait-to-fail" method for identifying students needing extra support (Lane et al., 2012). On the other hand, behavior screeners evidence strong psychometric properties, allow schools to get ahead of the referral process for all students, and are more successful at identifying students with internalizing issues. However, questions remain regarding the practicality and feasibility of conducting screeners. Not only do some screeners (e.g., SSBD) cost money to purchase, they also may be time-consuming to complete. Thus, schools must balance issues of practicality, efficiency, and effectiveness when determining how to identify students for Tier 2 supports (Lane et al., 2012). We encourage future researchers and practitioners to use multiple data sources with academic and behavioral information to identify and support students for whom the Tier 1 behavioral plan is not enough.

Student characteristics. Whereas authors of the reviewed studies were very clear about how students were identified for participation in Tier 2 supports, the characteristics of these students was slightly less lucid in terms of special education status. Given all students, regardless of disability status, have access to the Tier 1 plan, it is important to clarify the types of students who may be responsive to the Tier 1 plan and those who need Tier 2 supports. Additionally, we are interested in knowing whether students receiving Tier 2 supports who were not responsive (a) went on to receive Tier 3 interventions, or (b) were referred to a student support team for a special education evaluation. This information may lend evidence to the effectiveness of multitiered models in preventing at-risk students from developing more

severe problems meriting special education services and should be included in future research.

Also of interest is the screening status of participants. For example, on the SSBD, students with externalizing behaviors were identified almost three times as often as students with internalizing behaviors. An additional 219 participants were identified with risk, but the type was not specified. This highlights two important points—one, students with internalizing behaviors may go undetected and underserved (Cheney et al., 2008), and two, it is critical researchers report specific screening types to clarify for whom what intervention works best.

Finally, the students in this sample of studies lacked diversity in gender and ethnicity, as most were male and White, respectively. We question whether the limited sample diversity was a function of locale, given all studies were conducted on the west coast (most frequently in the Pacific Northwest) or the south (most frequently, central Tennessee) by a handful of researchers. We recommend examining this issue further by encouraging different researchers to conduct studies in a variety of locales to meet the call for establishing evidence-based practices (Horner et al., 2005). In a related way, the school level of participants was fairly homogeneous. The large majority of studies occurred at the elementary level, which may indicate a dearth of multitiered prevention models implemented at middle and high school levels. Because we know younger students are more amenable to intervention (Kazdin, 1987) and, clearly, targeted interventions are occurring at the elementary level, we need to find ways to effectively intervene with what is a presumably more resistant population at middle and high school levels.

Measurement

As we move toward establishing evidence-based Tier 2 interventions, experimentally rigorous designs are necessary to draw accurate conclusions about effectiveness. Although many of the included descriptive studies evidenced positive behavioral changes, reported the reliability of dependent variables, and measured and reported treatment integrity and social validity; descriptive designs are not sufficient for drawing causal conclusions about intervention effectiveness (Johnson, 2001). To this end, members of the field (i.e., Gersten et al., 2005; Horner et al., 2005) have outlined core indicators for conducting and evaluating group experimental and single-subject research. Granted it is beyond the scope of this article, future researchers should consider evaluating the included group experimental and single-subject studies for adherence to these indicators to establish specific Tier 2 interventions as evidence-based.

We did, however, examine several quality indicators such as dependent variable reliability, treatment integrity, and social validity. A particular strength of this body of literature was the reporting of dependent variable reliability, thus ensuring effects were not due to measurement error, but rather, the intervention (Horner et al., 2005). We were encouraged not only by the number of authors reporting reliability, but also by the high IOA occurring in all studies involving direct observation. This strength is tempered, however, by only one study reporting a kappa value, which is a more conservative estimate accounting for chance agreement (Watkins & Pacheco, 2001). And IOA, while frequently collected across all phases of studies, was often reported as an overall mean by participant and/or dependent variable(s) rather than for each phase as per recommended practices in single-subject research (Horner et al., 2005).

Another strength of the literature base was the number of studies reporting treatment integrity of the Tier 2 intervention, with reports indicating moderate to high integrity in terms of adherence to intervention components. Clearly, this has become an essential marker of scientific rigor as even authors failing to report Tier 2 integrity percentages astutely mentioned this omission as a serious limitation of the study (Todd et al., 2008). Moving forward, other dimensions of integrity like exposure, quality of delivery (which was addressed in only a few studies), participant responsiveness, and program differentiation may provide a more detailed account of differences between interventionists (Dane & Schneider, 1998). Like recommendations for Tier 1, a multi-method, multi-informant approach measuring instructional components and decision-making procedures associated with the intervention also may provide worthy information to researchers and practitioners alike. Although some studies included multiple perspectives, there is room for growth in this area.

Finally, because 10% to 15% of the student population will require Tier 2 supports, issues of practicality and feasibility must be addressed. Authors have met this call, as social validity assessments of a variety of stakeholders were prominent across studies, and results were generally favorable. One suggestion for improving this body of literature is to consider conducting cost-benefit analyses of interventions to determine the practicality, feasibility, and sustainability within specific schools, as resources vary by school.

Interventions and Associated Outcomes

Tier 2 interventions varied in terms of targeted domains (e.g., academic, behavioral, and/or social) and were evaluated using a range of experimental designs. The most commonly researched Tier 2 intervention in this body of literature included derivatives of CICO procedures (e.g., BEP, CCE, CICO). Although the experimental designs and details of the interventions varied slightly, the majority of participants in these studies showed decreases in problem behavior and increases in academic engagement. Despite

these positive effects, questions have been raised regarding the characteristics of students for whom these interventions are most effective. Because researchers have found function and severity of behavior may play a role in student responsiveness (e.g., Hawken et al., 2011; McIntosh et al., 2009), we encourage future researchers to address the extent to which these variables moderate responsiveness.

Further, we were surprised by the paucity of studies in the social domain given (a) many students with or at risk for EBD exhibit social skills deficits, and (b) social skills instruction embedded as a Tier 2 support within multitiered prevention models has been encouraged for remediating the behavioral problems of at-risk students (Walker & Severson, 2002). Nevertheless, we were compelled to mention a specific strength of the only study examining the effects of social skills instruction (Lane et al., 2003). In this study, the SSRS (Gresham & Elliott, 1990) was used to identify critical acquisition deficits that subsequently became the objectives for lessons derived from the *Social Skills Intervention Guide* (Elliott & Gresham, 1991). Using the rating scale and intervention guide together was particularly salient because it provided an efficient way to identify skill deficits and provide students with instruction tailored to meet these areas for improvement, and it resulted in positive student outcomes. As researchers and practitioners select Tier 2 supports, we encourage them to consider interventions that confidently reflect the skill deficits they are targeting.

Finally, academic skills were targeted in eight studies, with all studies demonstrating positive student outcomes. Some studies focused on only academic skills, whereas others included additional social or behavioral components. Surprisingly, initial studies of academic and behavioral components (i.e., SRSD plus reinforcement) did not include measures of behavior, despite participants having identified behavior problems (Lane et al., 2008; Lane et al., 2010; Little et al., 2010). The most current study of SRSD plus reinforcement, however, did include measures of behavior with students in the treatment group improving on writing probes and on academic engagement (Lane et al., 2011). In studies of literacy instruction (Lane et al., 2002; Wills et al., 2010), however, proximal measures of knowledge acquisition related to the curriculum and distal measures of behavior were included. These studies showed explicit, structured literacy instruction resulted in (a) more time for teachers to engage in instruction rather than behavioral management (Wills et al., 2010), and (b) improved reading and behavioral outcomes (Lane et al., 2002). By demonstrating the effects of academic interventions on behavior, researchers may help illuminate the direction of the relation between academic and behavior problems—a question that has been difficult for researchers to answer (McIntosh et al., 2012). Much more research is needed in this area, including studies that include measurement of academic and behavioral outcomes, and studies comparing the effects of academic-only, behavior-only, or combined, academic and behavioral interventions on students exhibiting both types of problems.

Limitations and Final Considerations

In sum, our findings revealed several limitations of the literature base. First, we identified a noteworthy gap in the research base for reporting Tier 1 treatment integrity in Tier 2 intervention studies, with the SET being the most frequently used assessment tool. Although the SET provides useful information about the school as a whole, it may be insufficient for assessing individual teachers' implementation of Tier 1 as it does not explicitly measure implementation at the classroom level, nor does it capture all Tier 1 efforts in a school (e.g., literacy programs and social skills programs). Therefore, we recommend multi-method, multi-informant measurement at the classroom level to more accurately determine which students are truly nonresponsive to Tier 1 instruction, and thus may need targeted intervention. The goal is to avoid false negatives (indicating a student does not need a Tier 2 support when in actuality he/she does) and false positives (indicating a student does need a Tier 2 support when in actuality he/she does not), with the former error being the most serious (Lane et al., 2012). Second, we examined participant characteristics and determined multiple data sources (e.g., ODRs, systematic behavior screeners, and academic measures) should be used to accurately identify and support students in Tier 2 interventions. Third, we found the participant sample lacked (a) diversity in terms of behavioral type (e.g., internalizing), gender, ethnicity, grade level, and locale; and (b) clarity in terms of special education status. Thus, we encourage future researchers to include more diverse samples and provide increased attention to students with disabilities and internalizing behaviors. Finally, the need for scientifically rigorous studies to establish evidence-based, Tier 2 interventions exists, as some studies in this review were limited by methodologies that prevented causal conclusions to be drawn.

In addition to limitations of the literature base, limitations of review procedures must be considered. For example, this review was limited by including only studies that documented a Tier 1 behavioral plan. As the field moves toward an integrated approach to multitiered models of prevention, we need to explore models that evidence academic, behavioral, and social components in the Tier 1 plan (Lane et al., 2012). In addition, studies were not evaluated according to all of the core quality indicators for educational research presented by Gersten et al. (2005) and Horner et al. (2005). Given the call for establishing evidence-based practices, we acknowledge the need to evaluate studies for scientific rigor using the quality indicators in a step towards this effort.

Despite the limitations of our approach to the review and the literature base as a whole, our findings indicated that many students, particularly those exhibiting externalizing behaviors, achieved varying degrees of success as a result of interventions in academic, behavioral, and social domains. Moving forward, we challenge researchers and practitioners to select interventions that address (a) the whole student, (b) specific skill deficits, and (c) various behaviors and functions. We also hope researchers will continue their efforts to meet the scientific standards set forth by Gersten et al. (2005) and Horner et al. (2005) when designing, implementing, and evaluating interventions conducted within multitiered models of prevention.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

*Campbell, A., & Anderson, C. M. (2008). Enhancing effects of check-in/check-out with function-based support. *Behavioral Disorders*, *33*, 233–245.

*Campbell, A., & Anderson, C. M. (2011). Check-in/check-out: A systematic evaluation of component analysis. *Journal of Applied Behavior Analysis*, *44*, 315–326. doi:org/10.1901/jaba.2011.44-315

Cartledge, G., Sentell, J., Loe, S., Lambert, M. C., & Reed, E. S. (2001). To be young, gifted, and black?: A case study of positive interventions within an inner-city classroom of African-American students. *Journal of Negro Education*, *70*, 243–254. doi:org/10.2307/3211277

Chafouleas, S. M., Riley-Tillman, T. C., Sassu, K. A., LaFrance, M. J., & Patwa, S. S. (2007). Daily behavior report cards: An investigation of the consistency of on-task data across raters and methods. *Journal of Positive Behavior Interventions*, *9*, 30–38. doi:org/10.1177/10983007070090010401

*Cheney, D., Flower, A., & Templeton, T. (2008). Applying response to intervention metrics in the social domain for students at risk of developing emotional or behavioral disorders. *Journal of Special Education*, *42*, 108–126. doi:org/10.1177/0022466907313349

*Cheney, D., Stage, S., Hawken, L., Lynass, L., Mielenz, C., & Waugh, M. (2009). A two-year outcome study of the check, connect, and expect intervention for students at-risk of severe behavior problems. *Journal of Emotional and Behavioral Disorders*, *17*, 226–243. doi:org/10.1177/1063426609339186

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23–45. doi:org/10.1016/S0272-7358(97)00043-3

Drummond, T. (1994). *The Student Risk Screening Scale (SRSS)*. Grants Pass, OR: Josephine County Mental Health Program.

Elliott, S., & Gresham, F. M. (1991). *Social skills intervention guide: Practical strategies for social skills training*. Circle Pines, MN: American Guidance Service.

*Fairbanks, S., Sugai, G., Guardino, D., & Lathrop, M. (2007). Response to intervention: Examining classroom behavior support in second grade. *Exceptional Children*, *73*, 288–310.

*Filter, K., McKenna, M., Benedict, E. A., Horner, R. H., Todd, A. W., & Watson, J. (2007). Check in/check out: A post-hoc evaluation of an efficient, Tier 2-level targeted intervention for reducing problem behaviors in schools. *Education & Treatment of Children*, *30*, 69–84.

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*, *60*, 294–309.

Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System (SRSS-T)*. Circle Pines, MN: American Guidance Service.

*Gresham, F. M., Van, M. B., & Cook, C. R. (2006). Social skills training for teaching replacement behaviors: Remediating acquisition deficits in at risk students. *Behavioral Disorders*, *32*, 363–377.

Hammill, D., & Larsen, S. (1996). *Test of Written Language–3*. Austin, TX: Pro-ED.

*Hawken, L. S. (2006). School psychologists as leaders in the implementation of a targeted intervention: The behavior education program. *School Psychology Quarterly*, *21*, 91–111. doi:org/10.1521/scpq.2006.21.1.91

*Hawken, L. S., & Horner, R. H. (2003). Evaluation of a targeted intervention within a schoolwide system of behavior support. *Journal of Behavioral Education*, *12*, 225–240. doi:org/10.1023/A:1025512411930

*Hawken, L. S., MacLeod, K. S., & Rawlings, L. (2007). Effects of the behavior education plan (BEP) on office discipline referrals of elementary school students. *Journal of Positive Behavior Interventions*, *8*, 94–101. doi:org/10.1177/10983007070090020601

*Hawken, L. S., O'Neil, R. E., & McLeod, K. S. (2011). An investigation of the impact of function of problem behavior on effectiveness of the behavior education program (BEP). *Education & Treatment of Children*, *34*, 551–574. doi:org/10.1353/etc.2011.0031

Hawken, L. S., Vincent, C. G., & Schumann, J. (2008). Response to intervention for social behavior: Challenges and opportunities. *Journal of Emotional and Behavioral Disorders*, *16*, 213–225. doi:org/10.1177/1063426608316018

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., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2004). The school-wide evaluation tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions*, *6*, 3–12. doi:org/10.1177/10983007040060010201

Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, *30*(2), 3–13. doi:org/10.3102/0013189X030002003

Kame'enui, E. J., & Simmons, D. C. (2003). *Planning and evaluation tool for effective school-wide reading programs*. Eugene, OR: Institute for the Development of Educational Achievement.

Kaminski, R., & Good, R. H., III. (1998). Assessing early literacy skills in a problem-solving model: Dynamic indicators of basic early literacy skills. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 113–142). New York, NY: Guilford.

Kamps, D., & Greenwood, C. (2005). Formulating Tier 2 level reading interventions. *Journal of Learning Disabilities*, *38*, 500–509. doi:org/10.1177/00222194050380060501

*Kamps, D., Wills, H. P., Heitzman-Powell, L., Laylin, J., Szoke, C., Petrillo, T., & Culey, A. (2011). Class-wide function-related intervention teams: Effects of group contingency programs in urban classrooms. *Journal of Positive Behavior Interventions*, *13*, 154–167. doi:org/10.1177/1098300711398935

Kauffman, J. M., & Landrum, T. J. (2009). *Characteristics of emotional and behavior disorders of children and youth* (9th ed.). Upper Saddle River, NJ: Merrill Prentice-Hall.

Kazdin, A. E. (1987). Treatment of anti-social behavior in children: Current status and future directions. *American Psychological Association*, *102*, 187–203.

Keller-Margulis, M. A. (2012). Fidelity of implementation framework: A critical need for response to intervention models of service delivery. *Psychology in the Schools*, *49*, 342–352. doi:org/10.1002/pits.21602

*Lane, K. L., Graham, S., Harris, K. R., Little, M. A., Sandmel, K., & Brindle, M. (2010). Story writing: The effects of self-regulated strategy development for second-grade students with writing and behavioral difficulties. *Journal of Special Education*, *44*, 107–128. doi:org/10.1177/0022466908331044

*Lane, K. L., Harris, K. R., Graham, S., Driscoll, S., Sandmel, K., Morphy, P.…Schatschneider, C. (2011). Self-regulated strategy development at tier 2 for second-grade students with writing and behavioral difficulties: A randomized control trial. *Journal of Research on Educational Effectiveness, 4*, 322-353. doi.org/10.1080/19345747.2011558987

*Lane, K. L., Harris, K. R., Graham, S., Weisenbach, J. L., Brindle, M., & Morphy, P. (2008). The effects of self-regulated strategy development on the writing performance of second-grade students with behavioral and writing difficulties. *Journal of Special Education*, *41*, 234–253. doi:org/10.1177/0022466907310370

*Lane, K. L., Kalberg, J. R., Mofield, E., Wehby, J., & Parks, R. J. (2009). Preparing students for college entrance exams: Findings of a Tier 2 intervention conducted within a three-tiered model of support. *Remedial and Special Education*, *30*, 3–18. doi:org/10.1177/0741932507314022

Lane, K. L., Menzies, H. M., Oakes, W. P., & Kalberg, J. R. (2012). *Systematic screenings of behavior to support instruction: From preschool to high school*. New York, NY: Guilford.

*Lane, K. L., Wehby, J. H., Menzies, H. M., Gregg, R. M., Doukas, G. L., & Munton, S. M. (2002). Early literacy instruction for first-grade students at-risk for antisocial behavior. *Education & Treatment of Children*, *25*, 438–458.

*Lane, K. L., Wehby, J. H., Menzies, H. M., Gregg, R. M., Doukas, G. L., & Munton, S. M. (2003). Social skills instruction for students at risk for antisocial behavior: The effects of small-group instruction. *Behavioral Disorders*, *28*, 229–248.

*Little, M. A., Lane, K. L., Harris, K. R., Graham, S., Story, M., & Sandmel, K. (2010). Self-regulated strategies development for persuasive writing in tandem with schoolwide positive behavior support: Effects for second-grade students with behavioral and writing difficulties. *Behavioral Disorders*, *35*, 157–179.

*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. doi:org/10.1177/10634266020100030401

Marchant, M., Anderson, D. H., Caldarella, P., Fisher, A., Young, B. J., & Young, K. R. (2009). Schoolwide screening and programs of positive behavior support: Informing universal interventions. *Preventing School Failure*, *53*, 131–444. doi:org/10.3200/PSFL.53.3.131-144

*Marchant, M. R., Solano, B. R., Fisher, A. K., Caldarella, P., Young, K. R., & Renshaw, T. L. (2007). Modifying socially withdrawn behavior: A playground intervention for students with internalizing behaviors. *Psychology in the Schools*, *44*, 779–793. doi:org/10.1002/pits.20265

*McIntosh, K., Campbell, A. L., Carter, D. R., & Dickey, C. R. (2009). Differential effects of a tier two behavior intervention based on function of problem behavior. *Journal of Positive Behavior Interventions*, *11*, 82–93. doi:org/10.1177/1098300708319127

McIntosh, K., Sadler, C., & Brown, J. A. (2012). Kindergarten reading skill level and change as risk factors for chronic problem behavior. *Journal of Positive Behavior Interventions*, *14*, 17–28. doi:org/10.1177/1098300711403153

Mitchell, B. S., Stormont, M., & Gage, N. (2011). Tier two interventions implemented within the context of a tiered prevention framework. *Behavioral Disorders*, *36*, 241–261.

*Mong, M. D., Johnson, K. N., & Mong, K. W. (2011). Effects of check-in/checkout on behavioral indices and mathematics generalization. *Behavioral Disorders*, *36*, 225–240.

Nelson, J. R., Martella, R. M., & Marchand-Martella, N. (2002). Maximizing student learning: The effects of a comprehensive school-based program for preventing problem behaviors. *Journal of Emotional and Behavioral Disorders*, *10*, 136–148. doi:org/10.1177/10634266020100030201

*Ness, B. M., Sohlberg, M. M., & Albin, R. W. (2011). Evaluation of a second-tier classroom-based assignment completion strategy for middle school students in a resource context. *Remedial and Special Education*, *32*, 406–416. doi:org/10.1177/0741932510362493

OSEP Technical Assistance Center on Positive Behavioral Supports and Interventions. Retrieved from www.pbis.org on March *1*, 2010.

Roach, A. T., & Elliott, S. N. (2008). Best practices in facilitating and evaluating intervention integrity. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 195–208). Bethesda, MD: National Association of School Psychologists.

*Robertson, E. J., & Lane, K. L. (2007). Supporting middle school students with academic and behavioral concerns: A methodological illustration for conducting Tier 2 interventions within three-tiered models of support. *Behavioral Disorders*, *33*, 5–22.

*Simonsen, B., Myers, D., & Briere, D. E. (2011). Comparing a behavioral check-in/check-out (CICO) intervention to standard practice in an urban middle school setting using an experimental group design. *Journal of Positive Behavior Interventions*, *13*, 31–48. doi:org/10.1177/1098300709359026

Sugai, G., Lewis-Palmer, T., Todd, A., & Horner, R. H. (2001). *School-wide evaluation tool*. Eugene: University of Oregon.

*Todd, A. W., Campbell, A. L., Meyer, G. G., & Horner, R. H. (2008). The effects of a targeted intervention to reduce problem behaviors: Elementary school implementation of check in-check out. *Journal of Positive Behavior Interventions*, *10*, 46–55. doi:org/10.1177/1098300707311369

Walker, H. M., & Severson, H. H. (1992). *Systematic Screening for Behavior Disorders (SSBD): User's guide and administration manual*. Longmont, CO: Sopris West.

Walker, H. M., & Severson, H. H. (2002). Developmental prevention of at risk outcomes for vulnerable antisocial children and youth. 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. 177–194). Needham, MA: Allyn & Bacon.

Watkins, M. W., & Pacheco, M. (2001). Interobserver agreement in behavioral research: Importance and calculation. *Journal of Behavioral Education*, *10*, 205–212. doi:org/10.1023/A:1012295615144

*Wills, H., Kamps, D., Abbott, M., Bannister, H., & Kaufman, J. (2010). Classroom observations and effects of reading interventions for students at risk for emotional and behavioral disorders. *Behavioral Disorders*, *35*, 103–119.