Exceptional Children

Vol. 78, No. 2, pp. 181-198.

Behavior Intervention for Students With Externalizing Behavior Problems: Primary-Level Standard Protocol

GREGORY J. BENNER
University of Washington

J. RON NELSON
University of Nebraska

ELIZABETH A. SANDERS

NICOLE C. RALSTON
University of Washington

ABSTRACT: *This article examined the efficacy of a primary-level, standard-protocol behavior intervention for students with externalizing behavioral disorders. Elementary schools were randomly assigned to treatment (behavior intervention) or control (business as usual) conditions, and K–3 students were screened for externalizing behavior risk status. The final sample included 7 treatment schools (*n *= 44 students) and 6 control schools (*n *= 26 students). Results of multilevel models showed that students with externalizing behavior in the treatment schools had significantly lower levels of problem behavior than those in the control schools. A positive but statistically nonsignificant treatment trend was observed for increased on-task behavior. No effects were observed for academic skills. The positive effects of the behavior intervention were smaller in schools serving higher proportions of students with low socioeconomic status and for students who had higher baseline levels of externalizing behavior. The discussion includes the results, practical importance, and limitations.*

Many U.S. schools are using multitiered or response to intervention (RTI) instructional models to improve the academic outcomes of their students. A recent national survey conducted by the American Association of School Administrators (2009) reveals that school use of such models continues to rise. In April 2009, 71% of schools indicated they were either piloting, in the process of districtwide implementation, or had multitiered or RTI instructional

models in district use, as compared to 44% in 2007. Schools are increasingly using these RTI models across all grade levels.

Similarly, schoolwide positive behavior intervention and support (SWPBIS) programs also use a continuum of behavior interventions that are consistent with the core principles of RTI. SWPBIS offers a continuum of interventions that are systematically applied to students, based on their demonstrated level of need, and addresses the role of the environment as it applies to the prevention and improvement of behavior difficulties. The continuum of interventions typically includes primary (provided to all students), secondary (supplemental intervention provided to students at some risk of experiencing behavior difficulties: Primary + Secondary), and tertiary (specialized and intensive intervention provided to students at high risk of or experiencing behavior difficulties: Primary + Tertiary). In this context, it is of interest to validate interventions and supports that can be used in SWPBIS models at the primary, secondary, and tertiary levels.

Schoolwide positive behavior intervention and support (SWPBIS) programs also use a continuum of behavior interventions that are consistent with the core principles of RTI.

The focus of this study is on a primary-level standard-protocol (i.e., well-defined, multicomponent) behavior intervention (Think Time Strategy; Nelson & Carr, 2000). The behavior intervention is a consequence-based classroom management strategy that is used in response to students' initial noncompliant response to teachers' request for them to stop a problem behavior (see description of procedures in "Methods" section). The conceptual framework for the behavior intervention is based on coercion theory (Patterson, 1982). Within coercion theory, the immediate effects of adults' attempts to stop the problem behaviors of children not only make the situation worse (in terms of persistence and escalation) but also play a key role in establishing ongoing coercive adult-child interactions. Research on student-teacher interactions in the classroom suggest that the same coercive interpersonal interactions occur between teachers and students who evince externalizing behavior (e.g., Nelson & Roberts, 2000). As such, the primary-level behavior intervention evaluated in this study is targeted primarily at students who evince externalizing behavior problems.

The standard-protocol behavior intervention was recognized as a promising intervention program by the U.S. Department of Safe and Drug Free Schools in 2001 and is included in a compilation of research-based classroom management strategies (Marzano, Marzano, & Pickering, 2003). The methodology of previous efficacy studies, however, do not meet current standards used by groups such as the Institute of Education Sciences What Works Clearinghouse (What Works Clearinghouse, 2008) and other organizations seeking to identify evidence-based interventions backed by "strong" evidence of effectiveness. The present randomized control trial, conducted in school settings, is designed to provide rigorous evidence regarding the efficacy of the behavior intervention. Randomized control trials that are well designed and implemented are considered the gold standard for evaluating the effectiveness of an intervention (What Works Clearinghouse, 2008).

PREVIOUS EFFICACY STUDIES

The behavior intervention in the present investigation arose out of a line of research assessing the effects of a comprehensive primary-level SWPBIS model on the problem behavior of students who tend to evince problem behavior (Nelson, Martella, & Galand, 1998; Nelson, Martella, & Marchand-Martella, 2002). The standard-protocol behavior intervention used in the present investigation was a component of the comprehensive SWPBIS model evaluated in these studies.

Two studies assessed the efficacy of the standard-protocol behavior intervention for responding to disruptive classroom behavior (Benner, Nelson, Smith, & Roberts 2002; Nelson, Martella, & Galand, 1998). Nelson, Martella, and colleagues used a continuous time series design over 4 years to assess the relative

effects of adjusting the ecological arrangements to promote positive behavior, as well as establishing, teaching, and reinforcing rules and routines with and without the behavior intervention on office discipline referrals in an elementary school (n = 594). Results indicated that following the implementation of the behavior intervention, the mean number of formal office disciplinary referrals for rule violations per academic quarter decreased an average of four (range = 1–7) referrals per grade level. The mean number of office disciplinary referrals for rights violations per academic quarter decreased an average of 35 (range = 22–46) per grade level.

Benner and colleagues (2002) examined the effects of the behavior intervention on verbal and physical aggression and on-task behavior of elementary-age students. These students were receiving special education services for emotional disturbance and were served in two self-contained special education classrooms. The results of the multiple-baseline across-classroom design indicated that the group (n = 18) of students with emotional disturbance, on average, demonstrated substantial decreases in the number of severe challenging behaviors. The mean weekly number of teacher reported problem behaviors exhibited by students per day in Classrooms A and B decreased from an average of 4.44 to .83 and 1.69 to .61, respectively. Similarly, on average, the percentage of time on task evinced by students in Classrooms A and B increased from an average of 33% to 62% and 54% to 63%, respectively. These finding suggests that the day-to-day classroom environment was less volatile and students were more task engaged following the implementation of the behavior intervention.

CURRENT STUDY

Despite the positive results of previous behavior intervention research, earlier study designs did not meet current standards for establishing strong evidence supporting global behavior interventions. The current study extended the earlier research on classroom behavioral intervention in several important ways. First, we used a randomized control trial design to assess the efficacy of the behavior intervention. Because the intervention is a universal (Tier 1) intervention, we randomly assigned schools, rather than children, to treatment and control conditions. Because schools were randomly assigned to experimental conditions, analyses must test the intervention at the school level; hence, we acknowledged at the outset that the study would be underpowered to detect small treatment effects (see "Limitations" for a complete discussion).

A second strength of the current study is that we measured treatment fidelity. Prior studies did not establish the extent to which the behavior intervention was implemented as prescribed.

Third, we used direct behavior-observation procedures to document the effects of the behavior intervention on negative and positive student behaviors, whereas previous research relied on office discipline referrals or teacher reports of the frequency of problem behavior.

Fourth, we examined the effects of the behavior intervention on the academic performance of students because, as already described, researchers have found evidence of a link between externalizing behavior and academic performance (Lassen, Steele, & Sailor, 2006; Nelson, Benner, Lane, & Smith, 2004).

Fifth, we assessed the extent to which baseline levels of problem behavior moderated the treatment effects. Other studies have found that baseline levels of performance moderate the efficacy of a wide range of interventions (Al Otaiba & Fuchs, 2002, 2006; Nelson, Benner, & Gonzalez, 2003).

Finally, we examined treatment effects in the context of school-level poverty because previous studies have shown that poverty has an adverse effect on student outcomes (Brooks-Gunn & Duncan, 1997; Caldas & Bankston, 1997; Coleman, 1966; Hogrebe & Tate, 2010). The three major research questions for the current study are as follows:

1. What are the effects of the behavior intervention on the negative (problem) and positive (on-task) behaviors of at-risk students?
2. What are the effects of the behavior intervention on the academic performance of at-risk students?

3. Do student baseline behaviors and school poverty level moderate response to the behavior intervention?

METHOD

Research Design

We used a randomized control study design to evaluate the efficacy of the primary-level standard-protocol behavior intervention with students who evince externalizing behavior disorders. Fourteen elementary schools in a diverse urban district in the Northwest United States were randomly assigned to either the treatment group (n = 7) or a business-as-usual control group (n = 7). Student screening, consenting, and pretesting occurred from October through February of the 2007–2008 academic year. Implementation of the behavior intervention for treatment schools began at the end of that year, in March 2008, and continued into the following 2008–2009 academic year through June 2009. This research obtained Institution Review Board (IRB) approval for all activities conducted.

Participants

Recruitment. All qualifying students (see Screening section below) in Grades K–3 at 14 diverse, urban, public elementary schools located in one school district in the Pacific Northwest were invited to participate. Schools were randomized into treatment and control conditions before screening, using a uniform random number generator in MS Excel software. One school, a control site, declined participation after randomization but before study onset. Hence, 13 schools comprise our final sample. Student enrollments at these schools ranged from 359 to 638 each, with enrollments averaging 456 (SD = 58.3) for treatment sites and 482 (SD = 82.8) for control sites. The percentage of students receiving free or reduced-price lunch was 55% (SD = 8.8%) for treatment schools and 52% (SD = 22.3%) for controls. The percentage of students receiving special education services for the treatment and control schools was 15% (SD = 3.9%) and 13% (SD = 4.7%), respectively. Finally, the percentage of students classified as transitional bilingual was 18% (SD = 5.5%) and 20% (SD = 9.5%) for the treatment and control schools, respectively. Chi-square tests showed no significant differences on these school-level characteristics between the treatment and control schools (all ps > .05).

Screening. All students in the 13 participating schools were initially screened for behavioral risk status in the fall of the academic year (before pretesting), with the exception of students who were already receiving special education services for emotional disturbance (of an externalizing nature or who had a formal diagnosis of an externalizing nature per the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* [DSM-IV; American Psychiatric Association, 2000]). These students who were already identified for services were automatically invited to participate in the study. Screening employed a modified version of the Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1990). At Stage I of the SSBD, teachers ranked 10 of their students who best evinced externalizing characteristics in their classroom. At Stage II, teachers completed a Critical Events Index (i.e., 33 items representing low-frequency, high-intensity behavior), an Adaptive Behavior Scale (i.e., nine items representing prosocial behavior), and a Maladaptive Behavior Scale (i.e., 10 items representing nonsocial behavior) on the top three students exhibiting externalizing behaviors from Stage I. Students were screened into the study if they received either (a) 5 or higher on the critical events scale, or (b) 1 or more critical events and simultaneously both (i) less than a 30 on the Adaptive scale and (ii) more than a 35 on the Maladaptive scale. From this information, 129 students were initially identified for study participation. According to the manual, the SSBD has been shown to have test-retest reliability coefficients ranging from 0.79 to 0.90 and internal consistency coefficients ranging from 0.82 to 0.94 (Walker & Severson, 1990). The SSBD has been validated through item, factorial, concurrent, discriminate, criterion-related, predictive, and construct-validation procedures.

Final Sample. Of the 129 students invited to participate, we received 97 affirmative parent consents. After attrition, the final sample comprised 70 children (n = 44 treatment, 26 control). Attrition included any student who moved within the

TABLE 1

Student Characteristics

	Treatment n = *44*		*Control* n = *26*		
Characteristic	n	*%*	n	*%*	χ^2 *(1)*
Male	37	(84.1)	21	(80.8)	0.13
Minority	25	(56.8)	10	(38.5)	2.20
English language learner	4	(9.1)	1	(3.8)	0.68
Special education	15	(34.1)	8	(30.8)	0.08
Existing behavior plan	15	(34.1)	10	(38.5)	0.14
Grades K–1	17	(39.6)	11	(42.3)	0.09
Grades 2–3	27	(61.4)	15	(57.7)	0.09

Note. All chi-square *p*-values > .10.

first two waves of observations (described subsequently); thereafter, any student who moved was still included in our statistical analyses but with their missing observations treated as missing at random. Across schools, there were 27 treatment teachers participating during the 2007–2008 academic year (ranging from one to seven teachers per school) and 16 control teachers (ranging from one to five teachers per school).

Student demographic characteristics are reported in Table 1. Mean treatment student age in fall 2008 was 99.86 months (Range = 66–122, *SD* = 14.78), which is approximately 8.32 years. Mean control student age in fall 2008 was 98.65 months (Range = 76–127, *SD* = 16.71), which is approximately 8.22 years. Chi-square tests showed no statistically significant differences in the demographic characteristics between treatment and control conditions on any status variable, including students receiving special education and behavior plan services.

Behavior Intervention

The behavior intervention in the current study included five components: (a) a precision request from the teacher (i.e., teacher uses a short verbal statement to encourage the child to exhibit on-task social behavior and does not use threats, ultimatums, warnings, or repeated request), (b) assigning the behavior intervention, (c) a reflective period for student to gain self-control (i.e., thinking time), (d) a behavior-debriefing process, and (e) student reentry to the classroom. Teachers taught and reviewed these components with their students using PowerPoint training slides provided by project staff.

Training began by reviewing classroom behavioral expectations and discussing the importance of self-control to maintain a safe and positive classroom. Then teachers explained how they would respond when students were having a difficult time with self-control, or behaving inappropriately. The teacher explained that the precision request would be used first and demonstrated how it would be used (see explanation that follows). The students and the teacher rehearsed the precision request and how it would be used in each classroom context (e.g., independent work, whole class discussion, transitions) and how the teachers would deliver it. The same "explain, model, lead, test" procedures were used to teach the remaining steps. The five components of the standard-protocol behavior intervention follow.

1. *Precision Request.* When a student exhibited problem behavior that teachers deem as problematic to the learning process (for the student or peers), teachers asked the student to comply with a precision request (e.g., "I need your attention, please").
2. *Assigning Behavior Intervention.* If a student did not comply, teachers directed the student to a designated colleague's classroom for a short reflective period. In the case of more severe forms of problem behavior (e.g., verbal or physical aggression), teachers simply

directed students to a designated classroom for a short reflective period. In either case, the verbal interaction of the teacher was nonconfrontational, limited, unemotional, and matter-of-fact.

3. *Reflective Period.* When directed, the student moved to and entered the designated classroom. The teacher used a pass to signal or prompt students to move independently to the designated classroom. The colleague teacher then directed the student to a designated desk. This desk was located in an area that was free from distractions and limited the ability of the student to engage the colleague teacher or other students.
4. *Debriefing Process.* After the colleague teacher had observed the student sitting in a calm manner during the reflective period at the desk, the colleague teacher approached the student to initiate a debriefing process. The debriefing was always conducted at the first opportunity after the student had been calm and gained self-control. The length of the reflective period was behavior-dependent (not time-dependent). The debriefing process then began when the colleague teacher asked the student to provide an objective verbal description of the behavior that initiated the behavior intervention. This question helped the colleague teacher to determine if the student had gained self-control and was ready to return to their regular classroom. After the student provided an acceptable (e.g., calm, objective) verbal description of the behavior, the colleague teacher gave the student a Behavior Debriefing Form to complete in writing. The form included three questions: (a) "What was your behavior?" (b) "What do you need to do differently when you go back to class?" and (c) "Can you do it [replacement behavior]?" The form was completed by the student independently unless the student was unable to do so (i.e., preliterate children or children with a language barrier). In such cases, the colleague teacher conducted a verbal debriefing and wrote the student's responses on the form. If the student was unresponsive or defiant to the question about the behavior that led to the use of the behavior intervention, the colleague teacher responded by saying, "I'll be back to you," and returned to their regular duties until another appropriate break arrived (and the student was still sitting in a calm manner). Throughout this process, the colleague teacher did not cajole the student, and avoided being drawn into a discussion with the student. Again, all interactions between the teacher and student were nonconfrontational, limited, unemotional, and matter-of-fact. After the Behavior Debriefing Form was completed, the teacher checked the form for completeness (i.e., inappropriate and replacement behaviors were stated in objective terms and the child indicated he or she was ready to go back to the classroom). If the form was completed correctly, the colleague teacher directed the student to return to their regular classroom with the completed form in hand.
5. *Classroom Re-entry.* The student then returned to the regular classroom and was welcomed back by the regular teacher. Before taking a seat, the student stood by the door and waited for their regular teacher's acknowledgment and appraisal of the completed Behavior Debriefing Form. If the form was accurate, the regular teacher helped the student to reengage in the current academic work. If the debriefing form was inaccurate, the student was directed back to the designated colleague's classroom to repeat the reflective period.

Staff Development Procedures

Principals, teachers, and staff at each school received ongoing training throughout the study by project staff. Educators participated in two 3-hour training sessions (January–February 2008) before program implementation (March 2008), and received a stipend as compensation for their time associated with participation in the project. At these training sessions, a four-step process was used to train educators to implement the behavior intervention, as follows:

1. Trainers provided educators with an overview of the theory, research base, rationale, and implementation format for the behavior intervention.
2. Trainers modeled and practiced the implementation activities with educators.
3. Simulated practice conditions were conducted to ensure that teachers obtained a high level of skill performance.
4. Trainers provided structured feedback to educators on how proficient they were during simulated practice conditions.

Following training, two behavior coaches supported implementation at the seven treatment schools. The implementation coach helped teachers by (a) explaining the behavior intervention to parents, target students, and peers; (b) securing the cooperation and consent of all parties to participate in the program's implementation; and (c) demonstrating and supporting teachers' implementation of the behavior intervention (e.g., problem-solving implementation issues).

Coaching was conducted in three ways. First, coaches offered a weekly, optional after-school meeting time in the staff lounge, once per week. The weekly problem solving allowed treatment teachers the opportunity to informally share successes and problem-solve with the coach and with peers. Second, coaches made actual visits to classrooms at convenient times for the teacher and the class. Coaches used the fidelity checklist (see next section) to provide feedback on strengths and areas for improvement. Third, coaches held whole-staff meetings, at which behavior coaches provided brief booster training to all staff. Often the booster training involved modeling a component of the behavioral intervention (e.g., precision request) and a brief rehearsal of the component (e.g., practice the precision request in small groups).

Intervention Fidelity

Project staff used a fidelity checklist to evaluate adherence to program implementation in the treatment schools and the extent to which components of the intervention were in use at control (business-as-usual) schools. The fidelity checklist targeted five components of the behavior intervention, as follows:

- Precision request.
- Assigning the behavior intervention.
- Reflective period.
- Debriefing process.
- Classroom re-entry.

The higher the fidelity score, the more the teacher adhered to the intervention as it was meant to be implemented. Project staff rated each of the teacher participants (i.e., teachers of student participants) in the treatment and control schools on the five behavior intervention components using a 5-point rating scale ranging from 1 (Never Implements) to 5 (Always Implements). Each observation yielded an overall total score and separate component scores for each of the five behavior intervention components.

Two researchers conducted fidelity of implementation observations on participating treatment teachers three times and control teachers twice. Interobserver reliability checks were conducted on 5% of fidelity observations, and the correlation between the two researchers was .94. Concurrent observations conducted for such reliability checks occurred during each round of data collection throughout the year. Although this strong interrater agreement was found, any item disagreements were resolved through discussion and sharing evidence for ratings. After discussion and evidence sharing on item disagreements, raters reached consensus on the most accurate teacher rating on the individual item and made additions to the scoring rules as necessary to decrease disagreements in future fidelity scoring.

The mean component fidelity scores for treatment teachers (out of 5 possible points) for precision request, assigning the behavior intervention, reflective period, debriefing process, and classroom re-entry were 3.52 (70%), 4.61 (92%), 4.75 (95%), 4.42 (88%), and 4.71 (94%), respectively. The mean component fidelity scores for control teachers for precision request, assigning the behavior intervention, reflective period, debriefing process, and classroom re-entry were 0.86 (17%), 0.08 (2%), 0.20 (4%), 0.08 (2%), and 0.24 (5%), respectively. The mean total fidelity score (out of 25 possible points) was 22.01 (88%)

for treatment teachers and 1.46 (6%) for control teachers. Taken together, these results indicate that teachers in the treatment schools implemented the behavior intervention with a relatively high degree of fidelity, whereas use of components in the behavior intervention by teachers in the control school was low. In addition, as Table 2 shows in the correlation matrix, the relationship between total intervention fidelity scores and outcomes was small. This relationship was expected given the high degree of implementation of the behavior intervention by teachers in the treatment schools.

Classroom Environments

Baseline classroom atmosphere was assessed using the Classroom Atmosphere Rating Scale (CARS; Wehby, Dodge, & Greenberg, 1993). Project staff used the CARS during classroom observations to assess the quality of the instructional environment for all teachers. Each classroom was observed for 30 min during math or reading instruction, and rated on a scale of 1 (Very High) to 5 (Very Low). Items included levels of student compliance during structured times, compliance during transitions, adherence to rules, cooperation, interest and engagement, on-task behavior, and the degree to which the environment was supportive of student behavior. There was no significant difference on the CARS between treatment (M = 79%, SD = 13%) and control teachers (M = 77%, SD = 13%, $t(11) = 0.42$, $p > .05$) at baseline.

Student Outcome Measures

Direct Observations of Student Behaviors. We used the Stage Observation System (SOS; Stage, 2007) to collect student behavior data. The SOS includes momentary and partial-interval time-sampling procedures during classroom instructional activities (e.g., mathematics, reading). Students were not observed during other activities, such as music, art, recess, lunch, arrival, or dismissal. The sampling procedure for the SOS involved a 15-min CD that signaled each 10-s interval with a prerecorded verbal prompt. For example, after 10 s, the voice on the CD said, "Observe 1-1." After exactly 10 more seconds, the prerecorded prompt states "Observe 1-2," and so forth. Observers used the prerecorded CD with a CD player and an earpiece so that only the observer could hear the prompts.

Momentary time sampling was used immediately after each prerecorded verbal prompt. First, the observer immediately noted whether the student was on task at the moment that the prompt occurred. Next, the observer watched for 10 s to determine any occurrence of the student performing any of seven coded categories of problem behavior (described in the section "Behaviors").

Observation Scheduling. Special observers (see later description) conducted eight waves of observations, with each wave involving four 15-min observation samplings that were averaged for each wave. Each of the four 15-min samplings for a given student was conducted on separate days within a 2-week period. No students were observed more than once per day, and multiple students were not observed during the same sampling. In other words, observation data was unique to each student.

Observations were scheduled in a quasi-random order whereby teachers were contacted to provide 1 week of notice that their student(s) would be observed during a given period. Thereafter (Waves 2–8), observations of students were staggered according to their initial dates of observation (to minimize variation in data due to gaps in scheduling). Wave 1 began in March 2008 (baseline), Wave 2 in April/May 2008, Wave 3 in September 2008, Wave 4 in October/November 2008, Wave 5 in November/December 2008, Wave 6 in January/February 2009, Wave 7 in February/March 2009, and Wave 8 in April/May of 2009.

Behaviors. The operational definitions for eight behaviors—one on-task (positive) and seven categories of problem (negative) behaviors—were as follows:

1. *On task*—the student being oriented toward the appropriate activity. For example, the student is oriented to his textbook, paper and pencil, or teacher. On task was also coded where the student was oriented to other students when the activity is a group activity. Otherwise, the behavior was considered off task.
2. *Talking*—the student talking with others about nonacademic topics.

TABLE 2

Intercorrelations

Measure	M	SD	N	*1.*	*2.*	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*	*10.*	*11.*	*12.*	*13.*	*14.*	*15.*	*16.*	*17.*	*18.*	*19.*	*20.*
1. Treatment status	0.63	0.49	70	—																			
2. School poverty	0.52	0.14	70	–.12	—																		
Percentage on-task behaviors																							
3. Observation 1 (Baseline)	51.76	12.53	70	.03	.19	—																	
4. Observation 2	62.22	13.71	69	.18	–.11	.05	—																
5. Observation 3	65.95	12.56	61	.02	**–.34**	–.04	.05	—															
6. Observation 4	71.64	14.97	58	.01	**–.33**	–.07	.04	**.40**	—														
7. Observation 5	73.63	16.55	58	**.29**	**–.35**	–.04	.19	.23	**.27**	—													
8. Observation 6	76.59	12.53	56	.10	**–.37**	.17	–.08	**.31**	.25	**.62**	—												
9. Observation 7	76.74	14.20	56	.12	**–.31**	–.06	–.11	.16	**.35**	**.37**	**.41**	—											
10. Observation 8	77.85	16.78	55	**.33**	**–.35**	.18	.02	**.43**	**.37**	**.44**	**.60**	**.54**	—										
No. problem behaviors																							
11. Observation 1 (baseline)	95.14	57.53	70	–.19	–.23	**–.60**	–.19	**.29**	.04	.18	–.07	.09	–.03	—									
12. Observation 2	69.77	41.47	69	–.17	–.04	–.09	**–.63**	.08	–.02	–.02	.11	.09	.06	**.44**	—								
13. Observation 3	43.10	31.01	61	.06	.17	.03	–.10	**–.62**	–.20	–.15	–.08	–.06	–.10	–.13	.03	—							
14. Observation 4	42.33	29.86	58	–.08	**.40**	.09	–.01	**–.46**	**–.81**	–.24	**–.34**	–.18	**–.42**	.05	.09	.20	—						
15. Observation 5	40.19	35.05	58	–.21	.17	.24	–.22	–.11	–.22	**–.85**	**–.35**	**–.26**	**–.32**	–.17	.19	.10	.26	—					
16. Observation 6	29.07	24.26	56	.04	**.32**	–.07	.06	**–.32**	–.15	**–.48**	**–.74**	–.15	**–.40**	.07	–.07	.15	**.36**	**.32**	—				
17. Observation 7	27.30	32.96	56	–.03	.22	–.09	.17	–.16	–.25	–.23	**–.38**	**–.79**	**–.63**	–.03	–.04	.02	.22	.23	.14	—			
18. Observation 8	27.07	24.89	55	–.22	.24	–.18	.11	**–.30**	**–.27**	–.21	**–.36**	**–.49**	**–.78**	.15	.00	.08	**.42**	.21	.24	**.63**	—		
Academic skills																							
19. Observation 8 only	96.52	17.47	56	–.18	–.10	.13	–.17	.15	.23	.11	**.31**	.18	.18	–.03	.06	–.08	**–.29**	–.03	**–.27**	–.04	–.16	—	
Treatment fidelity																							
20. Observations 3–8 Mean	0.90	0.02	44	—	**–.35**	–.26	–.14	.07	**.49**	.12	**.36**	**.37**	.20	–.01	.13	.28	**–.54**	–.19	**–.40**	–.30	–.20	.15	—

3. *Out-of-seat*—the student being out of his seat.
4. *Provoking*—derogative name-calling or performing physically threatening postures or gestures (i.e., shaking a fist at someone), or throwing objects to get someone's attention.
5. *Noise*—any audible noise produced by the student that is superfluous to the task at hand (e.g., humming, whistling, singing, rapping, snapping fingers, making popping sounds, or tapping pencils or feet).
6. *Aggression*—forceful contact with another person. Aggression is slapping, hitting, shoving, punching, and swinging arms at another person; swinging an object at another person; or throwing something at another person that could result in injury (e.g., throwing pencils or books). This also included verbally threatening behavior such as stating that the student will physically attack or threatens to physically assault another.
7. *Tantrum*—stomping feet, shoving books off desktops, or pushing chairs in a fit of temper.
8. *Refusal*—an overt behavioral negative response to a teacher's request.

The mean percentage of on-task behaviors during each of the four 15-min samplings was computed for each of the eight waves of observations. For problem behaviors, the mean number of problem behaviors across each of the four 15-min samplings for each of the eight observations was computed.

Observers and Reliability. In this study, four data collectors conducted observations. None of these observers were involved in any part of the treatment implementation. Training of the observers occurred with the measure's author before commencement of baseline data collection at Wave 1 of the study (March 2008). Training began with 2 days of intensive training, including simulated practice sessions in November 2007. Before onsite observations, observers completed supervised practice using the SOS on students with externalizing behaviors in comparable, but nonparticipating classrooms in schools outside the participating school district. Interobserver agreement was 90%. Any disagreements during practice situations were resolved through discussion and sharing of evidence for ratings. After discussion on item disagreements, observers, practice supervisors, and project staff reached consensus on the most accurate scoring of on-task and problem behavior in question, and made adjustments to operational definitions as necessary to decrease future disagreements between observers when observing in participating classrooms.

During the first wave of onsite data collection, interobserver reliability checks were conducted on 20% of observations weekly. After this, reliability checks occurred for 5% of observations. Interobserver agreement for on-site data collection during the study ranged from 79% to 100%, with a mean of 95.6% (SD = 4.2%).

Academic Skills. Project staff administered the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001) Academic Skills cluster to all students in March of 2008 (baseline) and May of 2009 (posttest). The WJ-III Academic Skills cluster is composed of the Letter-Word Identification, Spelling, and Calculation subtests. The reported test-retest reliability coefficients of the WJ-III Academic Skills cluster, Letter-Word Identification subtest, Spelling subtest, and Calculation subtest are .96, .96, .88, and .80, respectively (Woodcock et al. 2001).

Analytic Approach

We adopted multilevel modeling as our primary analytic tool. As compared with traditional unilevel methods (e.g., analysis of variance and multiple regression), the more complex analysis method accounts for dependencies among student scores due to school membership, allowing for valid inferences to be drawn about relationships between student outcomes and school-level predictors without violating the assumption of independence. Because the true unit of analysis for testing treatment effects in this study is schools, rather than students, experimental condition (treatment vs. control) is treated as a school-level variable in our models.

Initial Behavior Observation Models. For our two pretreatment behavior models (i.e., on task, problem behavior), we employed 2-level hierarchical models, in which students' initial observation scores (Level 1, n = 70) were nested within schools (Level 2, n = 13). In these models, we

tested for differences between experimental conditions using a school-level dummy coded variable (+1 = treatment, 0 = control) .

Final Behavior Observation Models. For our final observation analyses, we employed 3-level models in which observations (Level 1, n = 8 per student) were nested within students (Level 2), within schools (Level 3). In these models, we used effect coding (+1 = treatment, –1 = control) to appropriately create interaction terms. Further, we standardized our student-level baseline and school-level poverty predictors (i.e., converted to z-scores) for ease of results interpretation. Because there was no evidence supporting their exclusion, we included all covariate interaction terms to ensure model internal validity (functional form).

Although some research using growth models aims to describe developmental trajectories (e.g., with a focus on testing predictors of the linear or quadratic slope parameters), the growth modeling for the current study was used to obtain a robust estimate of the final observation data. Importantly, students with any missing data beyond the first two measurement occasions were able to be included in our observation models because multilevel software employs a full maximum likelihood approach to missing data (i.e., the variance-covariances of all available data are used to estimate model parameters, which is preferable to casewise deletion for variance estimates, as well as maintenance of statistical power; Schafer & Graham, 2002). Time was coded in real months, according to the mean interval between observation periods, with the intercept fixed at the final observation (–15 = initial observation to 0 = final observation). Although many growth models fix the intercept at the initial measurement occasion (baseline) or at a midpoint in the trajectory (to avoid collinearity), the choice of fixing the intercept is often a matter of convenience (if growth trajectories are the research focus), or it can serve purposefully to test an a priori research question. We fixed the intercept at the final measurement occasion to reflect the statistical tests appropriate to our research questions (i.e., testing treatment effects and treatment moderators on final outcomes); the linear and quadratic growth parameters do not substantively change when the intercept is fixed at the first or final occasion.

In preliminary modeling, we found that quadratic growth models (i.e., allowing for a bend in change over time) fit the data significantly better than linear models; hence we retained quadratic models for our observation analyses, and allowed both growth parameters to freely vary among students and schools. Additional model information can be obtained from the third author.

Academic Outcomes Model. We analyzed students' academic skills (measured at the end of the intervention study year) similar to the final behavior observation models; however, growth was not estimated (hence some students were excluded from analysis because of missing academic outcome data), nor was baseline included because it was only measured at the end of the study year.

In all multilevel analyses, we used the hierarchical linear modeling software program HLM (Raudenbush, Bryk, & Congdon, 2004); we used the Statistical Package for the Social Sciences (SPSS; SPSS Inc., 2006; 1989–2004) to compute descriptive statistics.

Effect Size Estimates. Effect size estimates were calculated as the model-estimated treatment slope coefficient divided by the approximate standard deviation; the approximate standard deviation is computed by multiplying the model-implied standard error with the square root of the number of schools in the study. The simple pooled standard deviation from the observed data was not used because it does not account for the nonindependence of students' scores within schools, nor does it account for missing data. We term these predicted effect size estimates as d^* for convenience—similar in interpretation to Cohen's d.

RESULTS

Table 3 shows observed baseline and end-of-implementation descriptive statistics for outcomes. For descriptive purposes, intercorrelations among variables included in statistical analyses are provided in Table 2 (the exception is treatment fidelity, which was not included in analyses but is reported for readers' interest). Although not included in Table 2, we note that none of the individual student demographic characteristics

TABLE 3

Observed Student Means and Standard Deviations

	Treatment n = *44*			*Control* n = *26*		
Measure	N	M	SD	N	M	SD
Percentage on-task behaviors						
Baseline	44	52.04	12.28	26	51.28	13.18
15-month	34	82.15	14.25	21	70.88	18.51
No. problem behaviors						
Baseline	44	86.75	52.42	26	109.35	63.84
15-month	34	22.85	22.69	21	33.90	27.26
Academic skills						
15-month	35	94.14	18.88	21	100.48	14.40

(shown previously in Table 1) correlated with observed or predicted behavioral outcomes.

Initial Behavior Observations

Multilevel analyses of initial observation scores revealed no statistically significant differences between experimental conditions baseline on task and problem behaviors (treatment slope *t*-test *p*s = .40 and .99, respectively).

Behavior and Academic Outcomes

On-Task Behavior Model. Table 4 shows results from our observation multilevel models for percentage of on-task and number of problem behaviors. Although treatment did not significantly predict percentage of on-task behavior, treatment schools did tend to have more on-task behavior than controls by 2.56% (p = .056), holding all other things constant (d^* = .61). In addition, statistically significant main effects for student baseline and school poverty on on-task behavior were evident. Holding other variables constant, schools with more impoverished students (school free and reduced-price lunch enrollment one standard deviation above average, Z = 1) were predicted to have 3.60% less on-task behavior at the end of the year, compared to the average school poverty level (Z = 0), and students who were more at risk (baseline Z = –1, one standard deviation below average) were predicted to average 0.41% fewer on-task behavior at the end of the year, compared with typical students in the study (Z = 0).

To better understand the interactions, we graphed model-implied means for each subgroup (see Figure 1). Specifically, we contrasted predicted values from schools with average poverty (Z = 0 free or reduced-price lunch enrollment) to those with higher poverty (Z = 1), and similarly, we contrasted students who were typical at baseline (Z = 0 baseline on-task behavior) to those who were more at risk at baseline (one standard deviation below average, Z = –1). In the first graph of Figure 1, holding baseline constant, we see that treatment effects were weaker for schools with greater poverty levels (treatment schools with typical poverty levels were predicted to have 2.56% more on-task behavior than control schools, whereas more impoverished treatment schools were predicted to have 0.80% fewer on-task behavior than control schools). In the second graph of Figure 1, we see a similar pattern: holding poverty constant, students who had typical baseline behavior were predicted to have 2.56% more on-task behavior than controls by the end

Although treatment did not significantly predict percentage of on-task behavior, treatment schools did tend to have more on-task behavior than controls by 2.56%.

TABLE 4

End-of-Implementation Multilevel Model Results for Behaviors

	% On-task Behaviors				*No. Problem Behaviors*			
Fixed Effect	*Coeff*	SE	df	t	*Coeff*	SE	df	T
Final Obs (Intercept)	77.85	2.20	9	35.32***	26.17	3.74	9	6.99***
Treatment	1.28	0.58	9	2.19	–2.55	0.71	9	–3.59**
School ZFRL	–3.60	0.59	9	–6.05***	4.77	0.75	9	6.37***
Treat*ZFRL	–1.68	0.59	9	–2.83*	1.63	0.75	9	2.18
ZBaseline	0.41	0.04	68	10.05***	0.42	0.04	68	9.81***
Base*treat	0.12	0.04	68	2.84**	–0.03	0.04	68	–0.77
Base*ZFRL	0.05	0.03	68	1.55	–0.02	0.03	68	–0.75
Base*treat*ZFRL	0.01	0.03	68	0.29	–0.08	0.03	68	–2.64**
Linear growth	0.22	0.44	12	0.49	0.73	1.17	12	0.63
Quadratic growth	–0.10	0.03	12	–3.21**	0.35	0.10	12	3.30**
Random Effect	*Var*	df		χ^2	*Var*	df		χ^2
Final Obs (Intercept)								
Students	99.61	44		89.15***	373.31	44		74.10**
Schools	15.32	9		16.54	67.45	9		9.92
Linear growth								
Students	1.60	45		49.21	19.15	45		46.45
Schools	0.17	12		16.15	6.44	12		16.11
Quadratic growth								
Students	0.01	45		47.69	0.18	45		65.24*
Schools	0.00	12		17.94	0.07	12		24.62*

Note. Obs = observation; ZFRL = standardized free-reduced lunch percentage; Treat = treatment; Base = baseline; ZBaseline = standardized baseline.

*$p \leq .05$. ** $p \leq .01$. *** $p \leq .001$.

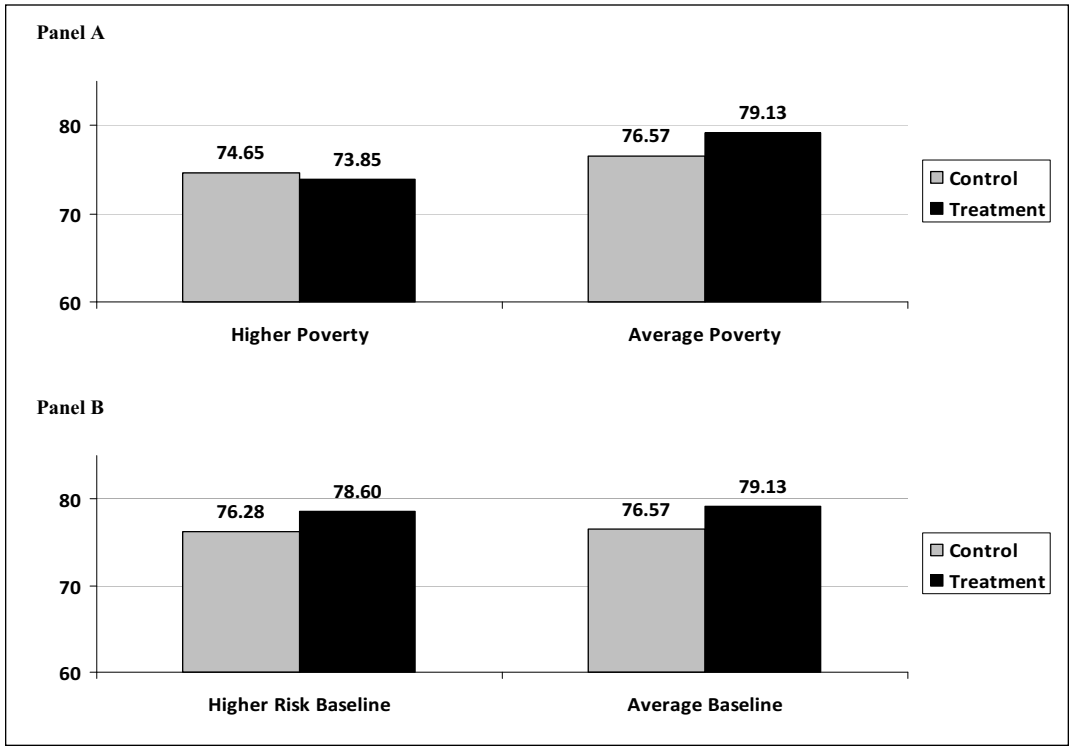
of the study, whereas the treatment effect for students starting out with lower baselines was only 2.32% more than controls.

Problem Behavior Model. Significant treatment effects were found for reducing problem behaviors, with treatment schools averaging 5.10 fewer problem behaviors at the final observation compared with the controls, holding all other things constant (d^* = –.99). We found that school poverty and baseline were significant treatment moderators, although student baseline was a moderator only when school poverty was taken into account (as evidenced by the significant three-way interaction). Analyses also showed that school poverty and student baseline were both significant unique predictors of problem behaviors. Holding other variables constant, schools with more impoverished students were predicted to have 4.77 more problem behaviors at the end of the study, compared to the average school poverty level, and students who were more at risk (baseline Z = 1, one standard deviation above average) were predicted to average 0.42 more problem behaviors at the end of the study, compared with typical students in the study.

To understand the interactions, we graphed model-implied means for each subgroup (see Figure 2). These graphs compare predicted values from schools with average poverty (Z = 0 free or reduced-price lunch enrollment) to those with higher poverty (Z = 1), juxtaposed with students who were typical at baseline (Z = 0 baseline problem behavior) to those who were more at risk at baseline (one standard deviation above average, Z

FIGURE 1

Model-Implied Values for End of Implementation: Percentage of On-Task Behavior by School Poverty Level (Panel A) and Student Baseline (Panel B)



= 1). Holding student baseline constant, treatment effects were weaker for schools with greater poverty levels (treatment schools with typical poverty levels were predicted to have 5.10 fewer problem behaviors than control schools, whereas more impoverished treatment schools were predicted to have only 1.84 fewer problem behaviors than control schools). When we inspected the differential performance among baseline values within school poverty level using predicted values, we found that treatment effects were fairly similar irrespective of student baseline levels in typical schools; however, in higher poverty schools, differential effects were predicted: students with higher risk baseline levels show 5.32 fewer problem behaviors compared with controls (similar to their counterparts in typical schools), whereas treatment students in impoverished schools who have typical baseline levels were predicted to have 1.84 fewer problem behaviors than controls.

Holding student baseline constant, treatment effects were weaker for schools with greater poverty levels.

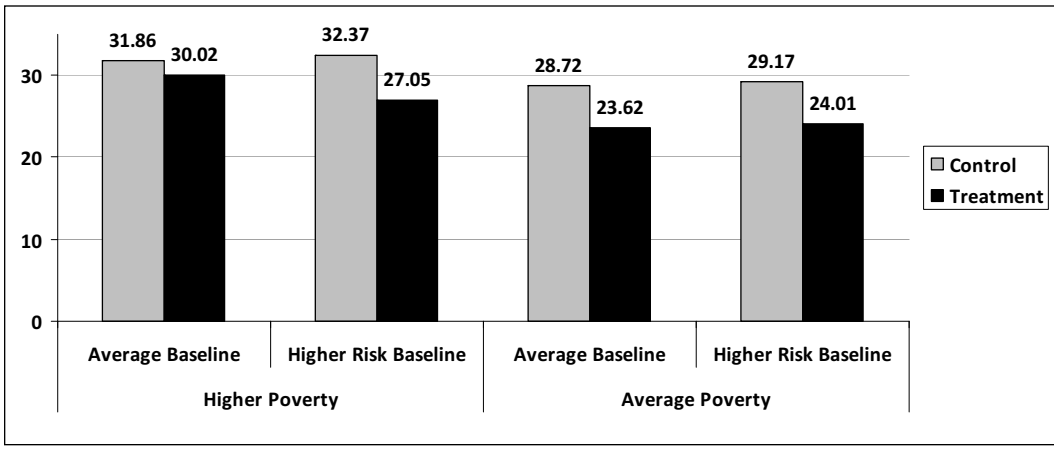
Academic Skills Model. Results of the multilevel academic skills model showed no significant main effect of treatment on WJ-III Academic Skills cluster scores (p = .07), no main effect of school poverty (p = .56), nor an interaction with school poverty (p = .23). Indeed, the academic performance of students in the treatment group was lower (d^* = –.58).

DISCUSSION

The primary purpose of this study was to test the efficacy of a primary-level behavior intervention designed for use by teachers as a response to the noncompliant behavior of students with external-

FIGURE 2

Model-Implied Values for End of Implementation: Number of Problem Behaviors by School Poverty Level and Student Baseline



izing behavior. The behavior intervention condition was compared to a business-as-usual control condition. Analysis results showed that the behavior intervention had a significant moderate effect on reducing problem behaviors (d^* = –.99) and a trend, albeit nonsignificant, for increasing percentage of on-task behavior (d^* = .61). Consistent with our model results, the zero-order simple correlations (of the observed data) in Table 2 show that problem behaviors were negatively correlated with percentage of on-task behaviors. Had our sample size been larger, we would have likely observed a significant main effect of treatment on on-task behavior. In addition, the findings of this study generally replicated the results of previous efficacy studies conducted on the behavior intervention (Benner et al., 2002; Nelson, Martella, et al., 1998).

These findings must be considered in light of the fact that the general socioeconomic level of the schools influenced the results. The benefits of the behavior intervention on behavioral outcomes were smaller in schools serving higher proportions of low socioeconomic students and for students who evinced higher baseline levels of externalizing behavior. This finding is consistent with the findings of a recent study of a three-tiered schoolwide positive behavior intervention and support (SWPBIS) model (Nelson, Duppong-Hurley, et al., 2009). The socioeconomic status of the schools moderated response to intervention on the behavior of students receiving secondary and tertiary level behavioral supports.

Our finding that treatment responsiveness was uniquely related to the general socioeconomic level of schools is also consistent with a plethora of research on the effects of school-level poverty on student outcomes (e.g., Brooks-Gunn & Duncan, 1997; Caldas & Bankston, 1997; Coleman, 1966; Hogrebe & Tate, 2010). This line of research suggests that school poverty concentration has an independent influence on education outcomes. More specifically, school poverty appears to have a limiting effect on both poor and nonpoor students alike (Hogrebe & Tate, 2010).

The relationship between school poverty concentration and student outcomes may be the result of several interacting peer, parent, and teacher variables. Students enrolled in high-poverty schools are more likely to encounter a peer atmosphere that is hostile to hard work, high achievement, and positive social behavior (e.g., Brooks-Gunn & Duncan, 1997). Further, parents of middle-class children tend to be more active in and more demanding of their children's schools (Hoover-Dempsy, Bassler, & Brissie, 1987). Finally, less qualified teachers are more likely to be teaching in high-poverty schools. In addition, teachers in these schools tend to transfer, resulting in high rates of teacher turnover, contributing further to the erosion in the delivery of

the curriculum and discipline practices (Caldas & Bankston, 1997).

As stated previously, we found that the effects of treatment were influenced by the students' initial baseline level of problem behavior. This finding suggests that the behavior intervention may be more effective with students with externalizing behavior who might be served at the secondary level of a three-tiered SWPBIS model. Our inclusion of students already identified and receiving special education services for emotional disturbance of an externalizing nature or who had a formal DSM-IV diagnosis of an externalizing nature, in tandem with our screening procedure used across all grades rather than just K–1 (as recommended to identify students at risk of behavior problems), resulted in a diverse sample of students who would typically be served in the secondary and tertiary levels of a three-tiered SWPBIS model. In general, this finding is not surprising. Students served at the tertiary level typically need intensive, individualized behavioral interventions to be successful (e.g., Horner, Sugai, Todd, & Lewis-Palmer, 2005; Nelson et al., 2009).

We found that the effects of treatment were influenced by the students' initial baseline level of problem behavior.

Although students who received the treatment showed improvements in their problem behavior relative to the control group, this was not the case in academic performance. These findings are inconsistent with previous studies that have found a small positive relationship between improvements in social behavior and academic performance (e.g., Lassen, Steele, & Sailor, 2006). This discrepancy may be a function of the fact that students' academic performance fell within the average range. In addition, the simple correlations among behavior outcomes and academic skills collapsed across both experimental conditions do not support the hypothesis that better student behavior is necessarily related to improved academic performance.

Practical Importance

It is difficult to fully assess the practical importance of the findings of the current study for several reasons. In our research, we have found no other randomized control trials conducted on primary grade level, consequence-based behavior interventions, approaches designed specifically for students evincing relatively high levels, or severe forms of externalizing behavior. Thus it is difficult to place our findings into an appropriate context. Quantitative effects of interventions are often described in terms of Cohen's (1988) *d*, and effect sizes of 0.20 are regarded "small" in magnitude, 0.50 "medium," and 0.80 "large." Based on more recent efforts to express intervention effects, researchers have indicated that effect sizes around 0.20 are of interest when they are based on primary-level interventions (Hedges & Hedberg, 2007) and when the research design uses a comparison group and randomization (Cohen, 1988). This view of research would suggest that the effect sizes for the current study warrant the interest of educators.

Limitations

Similar to most educational research, the present study has several limitations. Perhaps the most significant limitation is the location of the sample under study. The behavior intervention was studied in 13 elementary schools from one school district. Thus, the organizational and discipline structures, instructional practices, and demographic characteristics of the students and staff of the sample of schools limit the statements that can be generalized to schools in other settings. Although the schools in the current sample represent one of the more diverse districts in the region, the efficacy of the behavior intervention needs to be replicated with other diverse samples of schools.

The second limitation to this study is its lack of statistical power to detect treatment effects. We conducted a power analysis in Optimal Design for Multilevel and Longitudinal Research (Liu, Spybrook, Congdon, & Raudenbush, 2006) using the study's results as parameters (i.e., with 13 schools, an average of 5 students per school, and an intraclass correlation of .14 based on the variance component estimates given in the

results). We found that our power for detecting small, moderate, and large effects (d = .2, .5, and .8, respectively) was approximately 9%, 31%, and 65%, respectively.

Third, it is recommended that in future research, attempts be made to include a larger sample of schools. Of course, such studies will require extensive resources to collect data, given that school is the most appropriate unit of analysis for primary grade level interventions.

Fourth, attrition included an entire school (control site) that dropped from the study before the first observation wave; therefore we cannot be certain whether the results of this study would have changed had the school continued participation.

Fifth, the range in the grade levels of students participating in the study varied widely from kindergarten to Grade 3, which is a very broad period in child development. Future studies should investigate treatment efficacy by grade level to systematically test whether grade level moderates treatment effects.

Finally, the extent to which schools can implement the behavior intervention without the support of external resources is unclear. Teachers were provided ongoing coaching over the implementation period to ensure that the behavior intervention was implemented as prescribed (as appropriate for an efficacy trial). Future research on the behavior intervention should focus on implementation of these programs in school environments without extensive staff development support in an effectiveness study.

Despite these limitations, the current study adds to the body of evidence supporting use of a systematic, positive, schoolwide behavior program for schools and teachers who are currently lacking support in this area. Future research will shed light on the optimal contexts for implementing the program.

REFERENCES

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education, 23,* 300–316. doi:10.1177/07419325020230050501

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities, 39,* 414–431. doi:10.1177/00222194060390050401

American Association of School Administrators. (2009). *Response to intervention survey results.* Retrieved from http://www.spectrumk12.com/campaign/rti_survey_results

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, fourth edition.* Washington, DC: Author.

Benner, G. J., Nelson, J. R., Smith, D. J., & Roberts, M. L. (2002). The effects of a strategy to reduce the problem behaviors of students with emotional and behavioral disorders. *Academic Exchange Quarterly, 11,* 144–148.

Brooks-Gunn, J., &., & Duncan, G. J. (1997). The effects of poverty on children. *Future Child, 7,* 55–71. doi:10.2307/1602387

Caldas, S. J., & Bankston, C. L. (1997). The effect of school population socioeconomic status on individual student academic achievement. *Journal of Educational Research, 90,* 269–277.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Coleman, J. (1966). Equality and educational opportunity ["The Coleman Report"]. Washington, DC: U.S. Department of Health, Education, and Welfare.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group randomized trials in education. *Educational Evaluation and Policy Analysis, 29,* 60–87. doi:10.3102/0162373707299706

Hogrebe, M. C., & Tate, W. F. (2010). School composition and context factors that moderate and predict 10th-grade science proficiency. *Teachers College Record, 112,* 1008–1037.

Hoover-Dempsy, K. V., Bassler, O. C., & Brissie, J. S. (1987). Parent involvement: Contributions of teacher efficacy, socio-economic status, and other school characteristics. *American Educational Research Journal, 24,* 417– 435.

Horner, R. H., Sugai, G., Todd, A. W., & Lewis-Palmer, T. (2005) School-wide positive behavior support: An alternative approach to discipline in schools. In L. M. Bambara & L. Kern (Eds.), *Individualized supports for students with problem behaviors* (pp. 359–390). New York, NY: Guilford Press.

Lassen, S. R., Steele, M. M., & Sailor, W. (2006). The relationship of school-wide positive behavior support to

academic achievement in an urban middle school. *Psychology in the Schools, 43,* 701–712. doi:10.1002/pits.20177

Liu, X., Spybrook, J., Congdon, R., & Raudenbush, S. (2006). Optimal design software for multilevel and longitudinal research v1.77. Available at http://sitemaker.umich.edu/group-based/optimal_design_software.

Marzano, R. J., Marzano, J. S., & Pickering, D. J. (2003). *Classroom management that works: Research-based strategies for every teacher.* Alexandria, VA: Association for Supervision and Curriculum Development.

Nelson, J. R, Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta analytic review. *Learning Disabilities Research & Practice, 18,* 255–267. doi:10.1111/1540-5826.00080

Nelson, J. R., Benner, G. J., Lane, K., & Smith, B. (2004). Academic skills of K–12 students with emotional and behavioral disorders. *Exceptional Children, 71,* 59–74.

Nelson, J. R., & Carr, B. A. (2000). *The Think Time Strategy for schools.* Denver, CO: Sopris West.

Nelson, J. R., Duppong-Hurley, K., Synhorst, L., Epstein, M. H., Stage, S., & Buckley, J. (2009). The child outcomes of a behavior model. *Exceptional Children, 76,* 7–30.

Nelson, J. R., Martella, R., & Galand, B. (1998). The effects of teaching school expectations and establishing consistent consequences on formal office disciplinary actions. *Journal of Emotional and Behavioral Disorders, 6,* 153–161. doi:10.1177/106342669800600303

Nelson, J. R., Martella, R. C., & Marchand-Martella, N. E. (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:10.1177/10634266020100030201

Nelson, J. R., & Roberts, M. L. (2000). Ongoing reciprocal teacher-student interactions involving disruptive behaviors in general education classrooms. *Journal of Emotional and Behavioral Disorders, 8,* 27–37. doi:10.1177/106342660000800104

Patterson, G. (1982). *Coercive family process.* Eugene, OR: Castalia.

Raudenbush, S. W., Bryk, A., & Congdon, R. (2004). HLM6: Hierarchical linear and nonlinear modeling. Chicago, IL: Scientific Software International.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 72,* 147–177.

SPSS Inc. (2006). SPSS base 15.0 for Windows user's guide. Chicago IL: SPSS Inc.

Stage, S. (2007). Stage observation system (SOS-version V). Prepared for the Washington State Association of School Psychologists' 2007 Fall Convention in Spokane, Washington.

Walker, H., & Severson, H. (1990). Systematic screening for behavior disorders (SSBD). Longmont, CA: Sopris West.

Wehby, J. H., Dodge, K A., & Greenberg, M. (1993). *Classroom atmosphere rating scale.* Unpublished technical manual. Nashville, TN: Vanderbilt University.

What Works Clearinghouse. (2008). *WWC intervention report: Reducing behavior problems in the elementary school classroom.* Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practiceguides/behavior_pg_092308.pdf

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement.* Itasca, IL: Riverside.

ABOUT THE AUTHORS

GREGORY J. BENNER (Washington CEC), Professor, Education Program, University of Washington, Tacoma. **J. RON NELSON** (Nebraska CEC), Professor, Department of Special Education & Communication Disorders, University of Nebraska, Lincoln. **ELIZABETH A. SANDERS** (Washington CEC), Assistant Professor, Educational Psychology: Measurement, Statistics, & Research Design, University of Washington, Seattle. **NICOLE C. RALSTON** (Washington CEC), Research Analyst, Educational Psychology: Measurement, Statistics, & Research Design, University of Washington, Tacoma.

Address correspondence concerning this article to Gregory J. Benner, Education Program, University of Washington, Tacoma, 1900 Commerce Street, Box 358435, Tacoma, WA 98402 (e-mail: gbenner@u.washington.edu).

Manuscript submitted December 2010; accepted May 2011.