How Individual and School Aggregate Baseline Behavior Levels Moderate Response to a Primary Level Behavior Intervention

Gregory J. Benner
University of Washington, Tacoma

Elizabeth A. Sanders
University of Washington

J. Ron Nelson
University of Nebraska

Nicole C. Ralston
University of Washington

ABSTRACT: In our previous research (Benner, Nelson, Sanders, & Ralston, 2012), elementary schools were randomly assigned to either a primary-level behavior intervention directed at externalizing behavior (treatment, n = 7 schools) or to business-as-usual condition (control, n = 6 schools). A screening procedure was used to identify K through 3-grade students who exhibited externalizing behavior. The results showed that treatment students (n = 44) exhibited lower levels of problem behavior (ES = −0.99) and higher rates of on-task behavior (ES = 0.61) compared to controls (n = 26). The results also suggested that individual baseline behavior and school context as measured by free and reduced-priced meal rates moderated the findings. Given this latter finding, our interest in the current study was to test whether individual and school aggregate baseline behavior levels moderate students' response to two years of the behavior intervention. Schools originally assigned to treatment continued to implement the intervention during the subsequent academic year; schools in the control condition continued with their business-as-usual behavior management approaches. During the winter of the second year, students who remained in their schools were re-assessed on their behavior (n = 35 and n = 19 in treatment and control conditions, respectively). Analyses involving multilevel context modeling showed that students with relatively high initial baseline behavior levels (pretreatment) were more responsive to the intervention if they were in schools with higher school aggregate baseline behavior levels. Conversely, students tended to be less responsive to intervention if they were in schools with lower school aggregate baseline behavior levels. Limitations and practical implications are discussed.

■ The Conduct Problems Prevention Research Group (CPPRG, 1992; Dodge, Malone, Greenberg, & CPPRG, 2008) posited the dynamic cascade model of antisocial behavioral development from minor externalizing problems into serious antisocial behavior and violence depends on transaction with the environment across childhood. Environmental contexts and factors drive the development of antisocial behaviors in children whose behaviors emerge early in their lives. For example, fitting with the dynamic cascade model of antisocial behavior, the vital environmental factor of past and present coercive interactions with parents is predictive of the development of externalizing behavioral challenges. Often stemming from child biological difficulties and temperament, coercive family interaction patterns play a central role in the development of emotional and behavioral disorders (EBD) of an externalizing nature (Dodge et al., 2008; Patterson, 1982). Young children with biologically based difficulties in impulse control and behavioral regulation will likely have parents who experience difficulty with behavior management, particularly if the parents live in a

disadvantaged social-ecological context. The child's difficult temperament may grow into coercive interactions, and resulting behavior problems at home and impede the child from learning key school readiness skills (such as social-cognitive, language and vocabulary, and emotional skills) that emerge from healthy, positive, engaging interactions with caregivers.

Community researchers have identified naturally occurring factors (past and present) that influence individual mental health outcomes (Pickett & Pearl, 2001). Coinciding with the work of the dynamic cascade model, these researchers found that the effect of context over time and in current space predicts mental health outcomes. In their study of predictors of mental health in early adulthood, Wheaton and Clarke (2003) found that current neighborhood environment had no effect on early adult mental health, controlling for childhood neighborhood. Instead, earlier neighborhood settings have a lagged and cumulative effect on mental health differences. Thus, the history of neighborhood environments that the individual has lived in and moved through over time appears to be more important to adult mental health than current neighborhood environment. In the present investigation, we draw upon the analytical approach (see analytical approach in the Method section) used by these researchers to understand how the responsiveness of students with externalizing EBD is influenced in a similar manner by context stimuli when receiving behavioral intervention in the classroom (Marsh, Koller, & Baumert, 2001; Pickett & Pearl, 2001). As we will explain, our intent was to examine the school contextual factors (past and present) that contribute to student responsiveness to classroom behavioral intervention over two years.

Contextual Factors that Contribute to Responsiveness

Extant research on contextual factors that contribute to responsiveness to school-based behavioral intervention reveals that the socio-economic status (SES) level of a school moderates student outcomes (Brooks-Gunn & Duncan, 1997; Caldas & Bankston, 1997; Coleman, 1966; Hogrebe & Tate, 2010; Nelson et al., 2009). Socio-economic status represents sociological experiences that influence students' opportunity and likelihood to engage in certain behaviors. Thus, SES tends to represent the social ecology of schools (communities, families, etc.), which influences individual behavior, including the type, severity, and frequency (Hovell, Elder, Blanchard, & Sallis, 1986). Wiley, Siperstein, Forness, and Brigham (2010) found that lower school income was associated with higher levels of externalizing behaviors exhibited by individual students with EBD served in those schools. Baker, Kamphaus, Horne, and Windsor (2006) found that 17% of 1- through 5-grade students in low-income schools (70% free/reduced lunch) had pervasive, severe externalizing EBD, coupled with significant learning problems and limited self-regulation and social skills. Yet, these researchers found that only one third of these students were receiving special education services. In keeping with the dynamic cascade model of the development of antisocial behavior, higher severity and frequency of school aggregate disruptive behavior leads to high levels of teacher and administrator stress and burnout, and this is a significant reason why teachers leave the profession, which has an adverse effect on the achievement of students at risk of school failure (Berliner, 1986; Boyd et al., 2011; Espin & Yell, 1994; Harrell, Leavell, van Tassel, & Mckee, 2004; Ingersoll, 2001). Indeed, teachers cite behavior management concerns as their greatest challenge and are often shocked by the variety and intensity of behavioral issues they face (Meister & Melnick, 2003).

Researchers have found that student responsiveness to school-based behavioral interventions is shaped by student-level factors including baseline severity of behavior and chronicity of behavior (Gresham, 2005). For example, the impact of the Fast Track intervention on externalizing disorders was examined across childhood (CPPRG, 2011). Screened for risk of externalizing EBD at kindergarten, 891 children were randomly assigned by matched sets of schools to intervention or control conditions. Fast Track was implemented over 10 years and addressed parent behavior-management, child social-cognitive skills, reading, home visiting, mentoring, and classroom curricula. Significant interaction effects between intervention and initial risk level indicated that intervention prevented the lifetime prevalence of all psychiatric diagnoses (conduct disorder, oppositional defiant disorder, attention deficit hyperactivity disorder, and any externalizing disorder), but only among those at highest initial risk. High risk children screened in the

top third percentile have an 82% lifetime prevalence of an externalizing psychiatric disorder, compared with just 32% for the entire population. Indeed, the 10-year-long Fast Track intervention was found to prevent externalizing disorders over 12 years, including the 2-year period after intervention ended, among the highest risk group. Interestingly, the intervention did not have a positive effect among the group of children called "moderate risk," meaning children with behaviors between the 80th and 97th percentiles of risk.

School climate appears to impact student responsiveness to behavioral interventions. Researchers who have studied districts and schools implementing School-Wide Positive Behavioral Interventions and Supports (SWPBIS) have found that students in need of targeted (secondary level prevention) or intensive (tertiary level prevention) behavioral supports are most responsive when the overall school climate is safe and supportive (Lane, 2007; Lane, Oakes, & Menzies, 2010; Sugai, Horner, & Gresham, 2002; Walker, Ramsey, & Gresham, 2004). Stated differently, when primary level behavioral prevention systems are strong, likelihood of student responsiveness to secondary or tertiary level prevention efforts increases. Then SWPBIS sets the stage for student responsiveness to secondary and tertiary level behavioral interventions, given that the established host environments that support adoption and sustained use of evidence-based practices for students in need of secondary or tertiary behavioral supports are in place (Benner, Beaudoin, Chen, Davis, & Ralston, 2010; Lane, Kalberg, & Menzies, 2009; Zins & Ponte, 1990). The present study builds on this research by assessing how individual students with externalizing EBD tend to respond to school-based behavioral intervention by assessing whether individual student and school aggregate behavior levels moderate children's response to a primary level behavior intervention.

Original Study

In an earlier study (Benner, Nelson, Sanders, & Ralston, 2012), we conducted a randomized control trial on a primary level standard-protocol (i.e., well-defined, multi-component) 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 in the Method section below). 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 suggests that the same coercive interpersonal interactions occur between teachers and students who evince externalizing behavior (e.g., Nelson & Roberts, 2000; Shores, Gunter, Denny, & Jack, 1993). As such, the primary level behavior intervention evaluated in this study is targeted primarily at students who exhibit externalizing behavior problems.

A systematic screening process was used to identify K through 3-grade students who exhibited externalizing disorders. At posttest of the original study, treatment students (n = 44) exhibited fewer problem behavior (ES = −.99) and higher rates of on-task behavior (ES = .61) compared to controls (n = 26). Additionally, statistically significant main effects for student baseline and school poverty for on-task behavior were evident. Holding other variables constant, schools with more impoverished students were predicted to exhibit less on-task behavior at the end of the year compared to those students enrolled in schools with average school poverty levels, and students who exhibited lower rates of on-task behavior were predicted to lower rates of on-task behavior at the end of the year compared to those students who exhibited average rates of on-task behavior.

Current Study

The present study describes a 6-month follow-up of students who participated in the original study, allowing us to examine whether individual and school aggregate baseline behavior risk levels influenced their response to ongoing behavior intervention. We used established procedures (described below) to decompose student baseline behavior levels into uncorrelated between- and within-school components (e.g., Marsh et al., 2001; Wallace et al., 2007; Wheaton & Clarke, 2003). Schools originally assigned to treatment continued to implement the intervention during the subsequent academic year; schools in the control condition continued with their business-as-usual

behavior management approaches. During the winter of this second year, students who remained in their schools were re-assessed on behavioral outcomes.

Research Questions

For the current study, our guiding research questions were as follows.

1. *Main effects.* Do relative individual baseline and school aggregate student baseline behavior levels predict behavioral outcomes?
2. *Two-way interactions.* Does treatment responsiveness depend on relative individual and/or school aggregate student baseline behavior levels? Further, are students who exhibit relatively higher baseline behavior levels doing better or worse in schools with overall school aggregate behavior levels?
3. *Three-way interaction.* Is treatment responsiveness jointly dependent on both relative individual and school aggregate baseline behavior levels?

Method

Participants

The original study procedures and results are published in Benner et al. (2012). To briefly summarize, 13 urban, public elementary schools in the Pacific Northwest were randomized into treatment and control conditions prior to student screening. All students in the participating schools were initially screened for externalizing behavior in the fall of the academic year (before pretesting), with the exception of students who were already receiving special education services for emotional disturbance (these students were automatically invited to participate in the study). Screening employed a modified version (Stages I and II) 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 behavior 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. Of the 129 students invited to participate, we received 97 affirmative parent consents. After attrition, the sample comprised 70 children (n = 44 treatment, 26 control): 28 in grades K–1 (40%), 42 in grades 2–3 (60%), 58 males (83%), 35 minorities (50%), 5 English learners (7%), and 23 receiving special education services (33%). Treatment and control groups did not differ on any demographic characteristic.

Students who had participated in the original study were follow-up tested 6 months post-intervention (in the winter of the subsequent academic year). The treatment schools continued to implement the behavior intervention, and the control schools continued their business-as-usual procedures. Attrition due to students moving from their schools included nine treatment and seven control students. Hence, 54 students remained for the current study analyses (n = 35 treatment, 19 controls). Simple t- and chi-square tests showed no evidence for differences between original study participants (before attrition) and current study participants (after attrition) on any outcomes or demographic characteristics (all p-values $> .05$).

Behavioral Intervention

The behavior intervention includes 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. The five components of the standard-protocol behavioral intervention follow.

1. *Precision request.* When a student exhibited a problem behavior that the teacher deemed as problematic to the learning process (for the student or peers), the teacher asked the student to comply with

a precision request (e.g., "I need your attention please").

2. *Assigning behavior intervention.* If the student did not comply, the teacher 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), the teacher simply directed the student to a different designated classroom for a short reflective period. In either case, the verbal interaction of the teacher was non-confrontational, 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. The length of the reflective period was behavior dependent (not time dependent).
4. *Debriefing process.* Once the student had regained self-control during the reflective period and was sitting at the desk in a calm manner, the colleague teacher initiated a debriefing process. The debriefing process 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 gauge whether the student had gained self-control and was ready to return to their regular classroom. After an acceptable (e.g., calm, objective) verbal description of the behavior was provided, the student was given a Behavior Debriefing Form (BDF) 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., pre-literate 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 BDF. In the event that 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 his/her 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 non-confrontational, limited, unemotional, and matter-of-fact. After the BDF was completed, the teacher checked the form for completeness (i.e., inappropriate and replacement behavior 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 his or her regular classroom with the completed form in hand.
5. *Classroom re-entry.* The student then returned to his or her classroom, stood by the door, and waited for the classroom teacher's acknowledgement and appraisal of the completed BDF before taking his or her seat. If the form was accurate, the classroom teacher helped the student to reengage in academic work. If the debriefing form was inaccurate, the student was directed back to the designated colleague's classroom to repeat the reflective period.

Fidelity

A fidelity checklist was used by project staff to evaluate adherence to program implementation in the treatment schools. The checklist targeted five components of the behavior intervention described above. Project staff rated each of the teacher participants (i.e., teachers of student participants) in the schools three times during the academic year 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. The mean total fidelity score (out of 25 possible points) was 22.01 (88%). 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.

Behavior Measures

We used two types of behavior measures for this study: an observation-based measure of on-task and problem behavior in the classroom and a norm-referenced measure of externalizing behavior. Each of the three outcomes was measured prior to study onset (henceforth "baseline") and again approximately two years later (at the end of the two-year intervention period).

Observation Measure

The observation-based measure involved use of 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-minute CD that signaled each 10-second interval with a prerecorded verbal prompt. For example, after 10 seconds, 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 seconds to determine any occurrence of the student performing any of seven coded categories of problem behavior. The operational definitions for eight categories of behavior—one on-task (positive) and seven categories of problem (negative) behavior—were as follows:

1. *On-task.* The student was oriented toward the appropriate activity. For example, the student was oriented to his or her textbook, paper and pencil, or teacher. On-task was also coded where the student was oriented to other students when the activity was a group activity. Otherwise, the behavior was considered off-task.
2. *Talking.* The student was talking with others about nonacademic topics.
3. *Out-of-seat.* The student was out of his or her 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 was 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 was 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 threatening to physically assault another student.
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 behavior during each of the four 15-minute samplings was computed (higher scores indicate better performance). For problem behavior, the total number of problem behavior across each of the four 15-minute samplings was computed (higher scores indicate worse performance).

Each of the four 15-minute samplings for a given student was conducted on separate days (but for each classroom of participating students, within a 2-week period, to minimize classroom disruptions). 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 (schedules were adjusted occasionally to meet teacher schedules).

Four trained data collection assistants (none involved in treatment fidelity checks) conducted observations. Prior to on-site observations, interobserver agreement was established at 90% (all disagreements were resolved via evidence-based discussion). Further, on-site reliability checks by the last author occurred for an additional random 5% of observations and were found to be in 96% agreement.

Norm-Referenced Measure

We also measured student behavior using the Externalizing Composite scale of the *Behavior Assessment System for Children Teacher Rating Scale* (TRS: Reynolds & Kamphaus, 2004). We utilized the child version of the TRS, which measures adaptive and problem behavior for children ages 6 to 11. The TRS requires approximately 10–20 minutes to rate student behavior. Teachers rate the student's frequency of certain behavior on a 4-point scale ranging from "1" (Never) to "4" (Almost Always). The Externalizing Problems Composite scale measures disruptive behavior and is a composite of the following syndrome scales: hyperactivity, aggression, and conduct problems. Internal consistencies and test-retest reliabilities reported in the manual range from .80 to .90. Scores are reported as T-scores (adjusted for age), with higher scores indicating more problem behavior.

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. Additionally, in this study the unit of randomization was schools, rather than students; as such, treatment effects are appropriately tested at the school level.

To estimate the amount of variance school membership accounted for, we specified unconditional (intercept-only) models prior to the more complex models (intraclass correlations are reported in Results). To test our research questions, we specified context models similar to those used in other research that have tested individual and school aggregate connections to outcomes (e.g., Marsh et al., 2001; Wallace et al., 2007; Wheaton & Clarke, 2003) by decomposing student baseline scores into between- and within-school components (components are uncorrelated). To accomplish this, baseline (prior to study onset; two years ago) *student* behavior levels were group-mean (within-school) centered such that a student's baseline value on each measure was equal to the number of standard deviations the student was from their sample peers at their particular school (i.e., a relative-oriented z-score). Baseline *school* aggregate levels were computed as the mean of each school's student baseline behavior level (i.e., aggregated), and then, using these aggregate scores, scores were grand-mean (between-school) centered such that a school's behavior level value was in the number of standard deviations the school was from the other schools in the sample (i.e., z-scores among schools). Finally, condition was effect coded (+1 = treatment school, −1 = control school) in order to compute interaction terms among condition and individual and school aggregate baseline values. The general model was as follows.

$$\begin{aligned}\text{StudentOutcome}_{ij} = {} & \gamma_{00} + g_{01} * \text{SchoolCondition}_j \\ & + \gamma_{02} * \text{ZSchoolLevel}_j \\ & + \gamma_{03} * \text{SchoolCondition}_j \\ & * \text{SchoolLevel}_j \\ & + \gamma_{10} * \text{ZRelativeStudentLevel}_{ij} \\ & + \gamma_{11} * \text{SchoolCondition}_j \\ & * \text{ZRelativeStudentLevel}_{ij} \\ & + \gamma_{12} * \text{ZSchoolLevel}_j \\ & * \text{ZRelativeStudentLevel}_{ij} \\ & + \gamma_{13} * \text{ZSchoolCondition}_j \\ & * \text{ZRelativeStudentLevel}_{ij} \\ & * \text{ZSchoolLevel}_j \\ & + U_{0j} + r_{ij}\end{aligned}$$

In the model above, a student's final behavioral outcome is equal to the sum of: the conditional mean behavior score across all students adjusted for school membership (γ_{00}); the student's school's treatment effect on the student outcome (γ_{01}); the effect of the student's school's standardized aggregate behavior level as measured at baseline (γ_{02}); the effect of student's standardized baseline level relative to the other students at their school (γ_{10}); all interactions among these three main effects (γ_{03}, γ_{11}–γ_{13}); the deviation between the student's school's predicted value and the grand mean predicted value (between-school residual, U_{0j}); and the deviation between the student's predicted and observed value (within-school residual, r_{ij}). We assumed between- and within-school residual errors were normally distributed in

TABLE 1
Context Effects Model Results

	TRS Externalizing	# Problem Behavior	% On-Task Behavior
Fixed Effects	**Coefficient**	**Coefficient**	**Coefficient**
Intercept (conditional mean)	65.65***	12.30***	83.16***
Condition (1 = Treat, −1 = Contr)	−3.15	−0.74	6.18***
School baseline (SchBase)	0.00	1.65	5.80***
Cond*SchBase	−3.11	−1.98*	−8.47***
Student relative baseline (StudBase)	0.62***	−0.01	0.04
Cond*StudBase	0.25	0.03	0.07
SchBase*StudBase	0.31*	0.00	−0.07
Cond*SchBase*StudBase	0.47***	−0.06*	0.08
Random Effects	**Variance**	**Variance**	**Variance**
Between schools	29.02***	0.05	0.03
Residual	66.49	136.55	99.27

Note. N = 13 schools, 54 students. School Baseline = z-score of school aggregate baseline score; Student Relative Baseline = student z-score within school. Tests of student baseline and student baseline interactions based on df = 37; all other tests based on df = 9. For the first two outcomes, higher scores indicate worse performance; for the last outcome, higher scores indicate better performance.
*$p < .05$, **$p < .01$, ***$p < .001$.

all models. (We note that we examined use of a Poisson distribution with over dispersion for estimating the percent on-task outcome and the results did not substantively differ from assuming a normal distribution; hence for simplicity we report results assuming errors were distributed normally.) All context models were estimated using *HLM 7* (Raudenbush, Bryk, & Congdon, 1996–2011), and all descriptive statistics were estimated using *PASW/SPSS 18* (1993–2009).

Results

Preliminary results from our unconditional (intercept-only) model showed that intraclass correlations (ICCs: the percent of variance in outcomes accounted for by school membership) were <.01 for observed problem behavior, .18 for TRS externalizing behavior scale, and .22 for observed on-task behavior. Averaged across all three outcomes, about 13% of the variance in student behavior was accounted for by school membership. Observed bivariate correlations are provided in the Appendix (treatment students in upper diagonal and control students in lower diagonal).

Full model results (adjusted for school membership) are presented in Table 1. We discuss our model results below in terms of each of our three research questions.

Main Effects: Do relative baseline individual and school aggregate student baseline behavior levels predict behavioral outcomes?

The school aggregate baseline rate of on-task behavior was a uniquely positive predictive of on-task behavior outcomes. Schools with higher levels of school aggregate baseline rates of on-task behavior were predicted to have students who were more on task at the end of the study.

Two-Way Interactions. Does treatment responsiveness depend on relative individual and/or school aggregate student baseline behavior levels? Further, are students who exhibit relatively higher baseline behavior doing better or worse in schools with overall school aggregate behavior levels?

Negative two-way treatment interactions among condition and school aggregate baseline behavior levels were detected for both observation outcomes (on-task, problem be-

havior). These interactions were ordinal in that the treatment effect was observed only for schools with students with average or higher baseline problem and on-task behavior levels. Specifically, control schools that were one standard deviation above average on school aggregate baseline problem behavior were predicted to have students who exhibited 16.67 problem behaviors at the end of the study whereas treatment schools with the same school aggregate profile were predicted to have students who exhibited 11.23 problem behaviors (a difference of 5.44 fewer negative behaviors for treatment). Comparatively, schools with average school aggregate baseline problem behaviors were predicted to have a treatment-control difference of only 1.49 fewer problem behaviors for treatment, and schools with below-average school aggregate baseline problem behavior were predicted to have a treatment-control difference favoring controls by 2.47 fewer problem behaviors.

The pattern of results for on-task behavior is highly similar. Control schools that were one standard deviation below average on school aggregate baseline on-task behavior levels were predicted to have students who exhibited on-task behavior for 63% of observed intervals at the end of the study. In contrast, treatment schools with the same profile were predicted to have students who exhibited on-task behavior for 92% of observed intervals (a difference of 29% more on-task behavior for treatment). Comparatively, schools with average school aggregate baseline on-task behavior levels were predicted to have a treatment-control difference of only 12% more observed on-task behavior for treatment, and schools with above-average school aggregate baseline on-task behavior levels were predicted to have a treatment-control difference favoring controls by 6% more on-task behavior.

Further, are students who exhibit relatively higher baseline behavior levels doing better or worse in schools with overall school aggregate behavior levels?

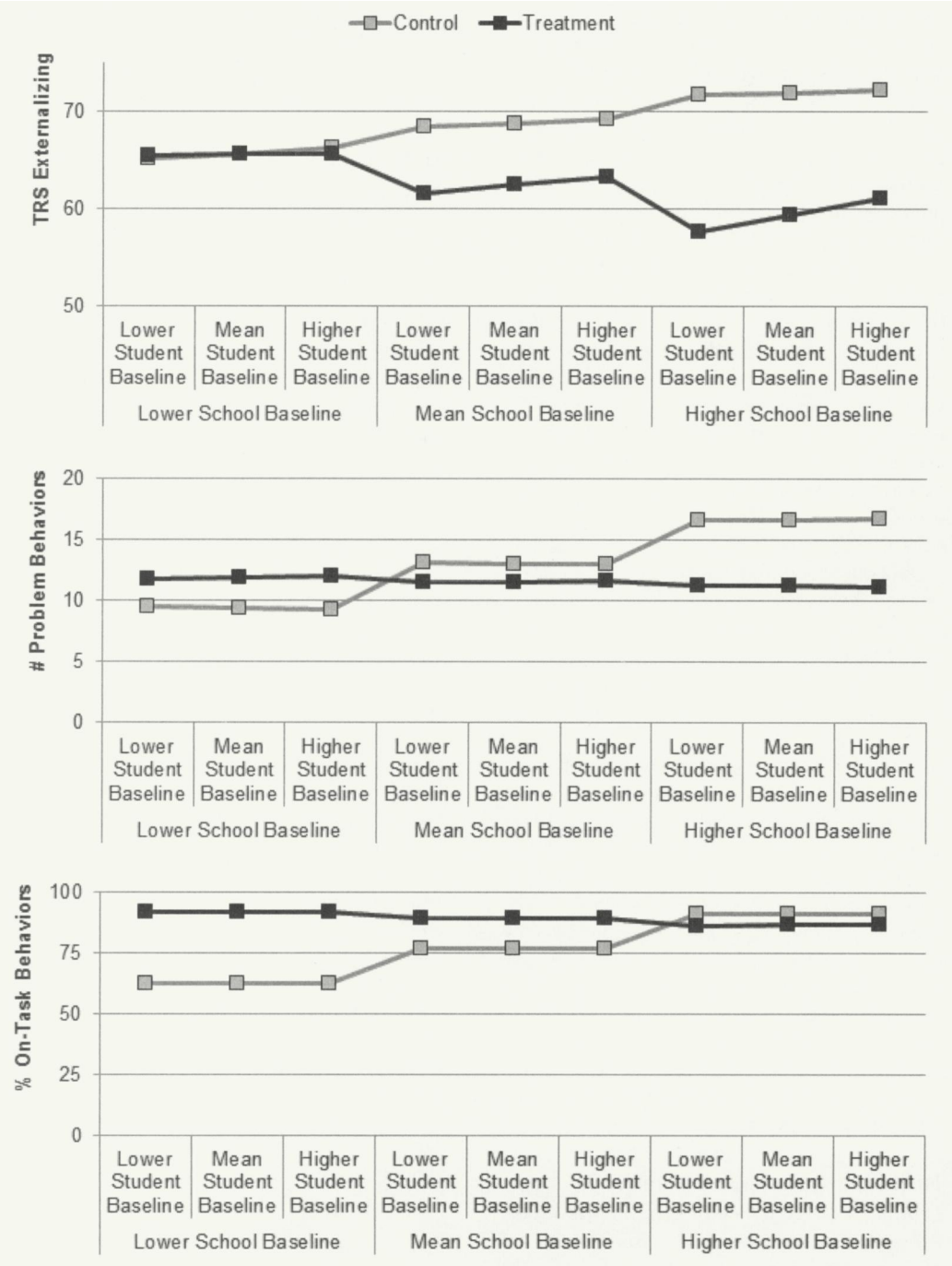
Significant two-way (ordinal) interactions between the school aggregate baseline and student relative baseline behavior levels were observed only for the TRS externalizing scale (but not the two observation outcomes). Holding treatment condition constant, students with lower relative baseline teacher ratings of externalizing behavior were predicted to have slightly better outcomes (lower scores at the end of the study), but only if they were in schools with higher school aggregate baseline levels (little advantage is predicted for the same students in schools with lower school aggregate baseline).

Three-Way Interaction. Is treatment responsiveness jointly dependent on both relative individual and school aggregate baseline behavior levels?

Model results detected three-way interactions for two of the three outcomes (each of these three relates to negative behavior outcomes). To understand the nature of these interactions, we computed predicted values for each condition (treatment and control) by three levels of our between- and within-school baseline levels (for simplicity: 1 *SD* below average, average, and 1 *SD* above average) and illustrate these in Figure 1. For reader interest, we computed these values for each outcome (although we note again that positive on-task behavior showed no significant three-way interaction). Inspection of the patterns in the predicted values shows that treatment benefits were estimated across all outcomes for students who were in schools with average or higher baseline behavior levels (as measured by school aggregate baseline values). However, students in schools with lower baseline behavior levels generally performed best if they were in a control (business-as-usual) school. For the TRS externalizing scale, students who also had lower baseline teacher ratings relative to peers in their schools (and who were in schools with higher school aggregate baseline levels) showed the largest treatment benefits; conversely, for the observed number of observed problem behaviors, students who had relatively higher baseline behavior levels and who were in schools with higher school aggregate baseline levels tended to have the greatest treatment benefits.

Discussion

The dynamic cascade model of antisocial behavioral development posits that growth from minor externalizing problems into serious antisocial behavior and violence depends on transaction with the environment across childhood (CPPRG, 1992; Dodge et al., 2008). Within schools, researchers have found that student responsiveness to school-based behavioral interventions is shaped by student level factors including baseline severity of behavior



Note. For first two measures, higher scores indicate worse performance; for last measure, higher scores indicate better performance.

Figure 1. Model-predicted (fitted) values by condition and baseline behavior risk levels. **Note. *For first two measures, higher scores indicate worse performance; for last measure, higher scores indicate better performance.***

and chronicity of behavior (CPPRG, 2011; Gresham, 2005). Researchers have also found that students with externalizing EBD who require intensive behavioral supports are most responsive with primary level SWPBIS prevention systems in place in a positive school climate whereby all students are actively taught behavioral expectations and acknowledged for appropriate behavior (Benner et al., 2010; Lane, 2007; Lane et al., 2010; Sugai et al., 2002; Walker et al., 2004; Zins & Ponte, 1990). The present study builds on this research by assessing the extent to which individual students with externalizing EBD are influenced in a similar manner by context stimuli (Marsh et al., 2001; Pickett & Pearl, 2001). More specifically, the current study extends previous research on the influence of school-level SES on student outcomes (Authors, 2012; Brooks-Gunn & Duncan, 1997; Hogrebe& Tate, 2010; Nelson et al., 2009) by testing whether individual and school aggregate behavior levels influence treatment response.

The primary level behavior intervention benefited students in both reducing negative behavior (classroom problem behavior and teacher ratings of externalizing problems) and improving on-task behavior. These treatment effects were stronger for students in schools with higher (more at-risk) school aggregate baseline behavior levels, and for students with relatively higher (more at-risk) behavior levels (compared to school peers). Control schools that were one standard deviation below average on school aggregate baseline on-task behavior levels were predicted to have students who exhibited on-task behavior for 63% of observed intervals at the end of the study. In contrast, treatment schools with the same profile were predicted to have students who exhibited on-task behavior for 92% of intervals (a difference of 29% more on-task behavior for treatment). Moreover, control schools that were one standard deviation above average on school aggregate baseline problem behavior were predicted to have students who exhibited 16.67 problem behaviors at the end of the study whereas treatment schools with the same school aggregate profile were predicted to have students who exhibited 11.23 problem behaviors (a difference of 5.44 fewer negative behaviors for treatment).

Students with higher risk baseline behavior levels responded best to primary level standard protocol behavioral interventions. Indeed, students with high levels of teacher-rated externalizing behavior, measured by the TRS, relative to peers in their schools (and who were in schools with higher school aggregate baseline levels) showed the largest treatment benefits. This finding coincides with those of the Conduct Problems Prevention Research Group (CPPRG, 2008). These researchers found that the Fast Track intervention prevented the lifetime prevalence of all psychiatric diagnoses but only among those at highest baseline risk over 12 years, including the 2-year period after intervention ended. These researchers also found that the effects of the intervention were not significant for students at moderate risk. In contrast, as discussed in the introduction, researchers have also found that baseline severity and chronicity of problem behavior reduces the likelihood of responsiveness to school based behavioral interventions (Gresham, 2005). Baseline levels of performance have been found to moderate the efficacy of a wide range of interventions (Al Otaiba & Fuchs, 2002, 2006; Nelson, Benner, & Gonzalez, 2003).

Our results show support for the behavior intervention across the middle and upper range of school aggregate behavior levels when behavior problems are defined by the TRS, which is a norm-referenced teacher rating scale. Further, qualified support for the behavior intervention was evident for schools with relatively higher school aggregate behavior levels as measured by the SOS, which is an observation measure. These somewhat contrasting findings may relate to the type of measurement of each of the outcomes, because the TRS is measured by teachers, whereas the SOS is measured by researchers. A plausible explanation is that the TRS captures nuances in children's behavior that the teacher observes over a longer period of time—nuances that the SOS cannot because it is a time sampling method of evaluating behavior. This said, it is important to note that the pattern of effects for SOS shows significant treatment interactions. According to our findings for the SOS, treatment was effective under certain conditions. Schools with higher school aggregate baseline levels tended to have children who were more responsive to the behavior intervention.

The finding that schools with higher school aggregate baseline levels (measured 2 years earlier) tended to have children who were more responsive to the behavior intervention was not expected. Nelson et al. (2009) reported that treatment effects for students with EBD who

received MultiSystemic Therapy (MST) within the three-tiered intervention program were moderated by the SES of schools (Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 1998). School poverty moderated treatment outcomes after 1 year of treatment. In the present investigation, higher school aggregate baseline behavior moderated treatment effects after 2 years of treatment. While current school behavior, climate, and safety are important to the responsiveness of students to behavioral interventions, baseline aggregate school behavior moderated treatment outcomes for individual students with EBD.

Limitations

As with all studies, this study has limitations. First, our sample size is relatively small—both in terms of schools (n = 13) and students (n = 54). Second, the results from this study can only be generalized to students who exhibit problem behavior at study onset (rather than all students), and to highly diverse schools (the district in which the schools were located is one of the most diverse districts in terms of ethnicity and language in Washington state). Third, attrition during the second year of the study was fairly high (33%); however, we tested whether those who moved differed from those who remained and found no evidence of differences on any characteristic or behavior at baseline. Fourth, the sample size is relatively small for detecting effects, and the statistically non-significant interactions may have been influenced by the smaller sample size. Finally, the length of the intervention was school-wide over a 2-year period; more research may be needed to examine how school-level change over a longer period of time affects student changes.

Practical Implications

Despite the limitations, the results of the current study suggest that the effectiveness of behavioral interventions will vary depending upon the individual and school aggregate behavior levels of students. Schools implementing the same behavioral intervention may achieve varied outcomes not only as a function of the quality of implementation, but also as a function of the social context of the school. Future research on School-Wide Positive Behavioral Interventions and Supports (SWPBIS) should explore the extent that creating a host environment influences the social environmental effects of baseline and school aggregate behavior levels. Not only may SWPBIS serve as a host environment that supports adoption and sustained use of evidence-based practices, but it may exert an influence on the social norms of students receiving behavioral interventions in the school. This social influence may interact with the quality of implementation to affect treatment outcomes. Finally, the results suggest that behavioral intervention research should attempt to account for the collective social influence of the population of students receiving treatments. Doing so may lead to implementation procedures to both enhance and inhibit the collective social influence of the treatment populations.

In addition, there are practical benefits of this investigation for elementary educators serving students with externalizing EBD, particularly those in schools serving students who exhibit severe and chronic behaviors. Teachers often struggle to maintain healthy and positive interactions with students with EBD, especially those with severe disruptive behaviors and limited academic engagement. Teachers often give multiple prompts, warnings, and ultimatums to respond to the behavior of these children. This response to problem behavior inadvertently makes the situation worse (in terms of persistence and escalation), makes the problem behavior functional, and contributes to ongoing coercive adult-child interactions (Nelson & Roberts, 2000; Shores et al., 1993). We suggest that the behavioral intervention be considered as a standard protocol approach for responding to the behavior of elementary students with externalizing EBD in the classroom in a calm, healthy, and therapeutic manner.

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.

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*(5), 414–431.

Baker, J. A., Kamphaus, R. W., Horne, A. M., & Windsor, A. P. (2006). Evidence for population-based perspectives on children's behavioral adjustment and needs for service delivery in schools. *School Psychology Review, 35*, 31–46.

Benner, G. J., Beaudoin, K. M., Chen, P. Y., Davis, C., & Ralston, N. C. (2010). The impact of intensive positive behavioral supports on the behavioral functioning of students with emotional disturbance: How much does fidelity

matter? *Journal of Behavioral Assessment and Interventions for Children, 1*, 85–100.

Benner, G. J., Nelson, J. R., Sanders, E. A., & Ralston, N. C. (2012). Efficacy of a primary level standard protocol behavior intervention for students with externalizing behavior problems. *Exceptional Children, 78*(2), 181–198.

Berliner, D. C. (1986). In pursuit of the expert pedagogue. *Educational Researcher, 15*(7), 5–13.

Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal, 48*(2), 303–333.

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

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.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, F., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). Equality and Educational Opportunity. Washington DC: US Government Printing Office.

Conduct Problems Prevention Research Group (CPPRG). (1992). A developmental and clinical model for the prevention of conduct disorders: The Fast Track program. *Development and Psychopathology, 4*, 509–527.

Conduct Problems Prevention Research Group (CPPRG). (2011). The effects of the Fast Track preventive intervention on the development of Conduct Disorder across childhood. *Child Development, 82*, 331–345.

Dodge, K. A., Malone, P. S., Greenberg, M. T., the Conduct Problems Prevention Research Group (CPPRG). (2008). Testing an Idealized Dynamic Cascade Model of the development of serious violence in adolescence. *Child Development, 97*, 1907–1027.

Espin, C. A., & Yell, M. L. (1994). Critical indicators of effective teaching for preservice teachers: Relationships between teaching behaviors and ratings of effectiveness. *Teacher Education and Special Education, 17*, 154–169.

Gresham, F. M. (2005). Response to intervention: An alternative means of identifying students as emotionally disturbed. *Education and Treatment of Children, 28*(4), 328–345.

Harrell, P., Leavell, A., van Tassell, F., & McKee, K. (2004). No teacher left behind: Results of a five-year study of teacher attrition. *Action in Teacher Education, 26*, 47–59.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (1998). *Multisystemic treatment of antisocial behavior in children and adolescents.* New York: Guilford.

Hogrebe, M. C., & Tate, W. F. (2010). School composition and context factors that moderate and predict 10th-grade science proficiency. *Teachers College Record, 112*(4), 1096–1136.

Hovell, M. F., Elder, J. P., Blanchard, J., & Sallis, J. F. (1986). Behavior analysis and public health perspectives: Combining paradigms to effect prevention. *Education and Treatment of Children, 9*, 287–306.

Ingersoll, R. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal, 38*(3), 499–534.

Lane, K. L. (2007). Identifying and supporting students at risk for emotional and behavioral disorders within multi-level models: Data driven approaches to conducting secondary interventions with an academic emphasis. *Education and Treatment of Children, 30*, 135–164.

Lane, K. L., Kalberg, J. R., & Menzies, H. M. (2009). *Developing schoolwide programs to prevent and manage problem behaviors: A step-by-step approach.* New York: Guilford.

Lane, K. L., Oakes, W. P., & Menzies, H. M. (2010). Systematic screenings to prevent the development of learning and behavior problems: Considerations for practitioners, researchers, and policy makers. *Journal of Disabilities Policy Studies, 21*, 160–172.

Marsh, H. W., Koller, O., & Baumert, J. (2001). Reunification of East and West German school systems: Longitudinal multilevel modeling study of the Big-Fish-Little-Pond Effect on academic self-concept. *American Educational Research Journal, 38*(2), 321–350.

Meister, D. G., & Melnick, S. A. (2003). National new teacher study: Beginning teachers' concerns. *Action in Teacher Education, 24*(4), 87–94.

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 and Practice, 18*(4), 255–267.

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*(1), 7–30.

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.

PASW/SPSS, Inc. (1989). *PASW Statistics 18.* Armonk, NY: IBM.

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

Pickett, K. E., & Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health, 55*, 111–122.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (1996–2011). *HLM for Windows 7.0.* Lincolnwood, IL: Scientific Software International, Inc.

Reynolds, C. R., & Kamphaus, R. W. (2004). *BASC-2: Behavior Assessment System for Children, Second Edition.* Circle Pines, MN: American Guidance Service.

Shores, R. E., Gunter, P. L., Denny, R. K., & Jack, S. L. (1993). Classroom influences on aggressive and disruptive behaviors of students with emotional and behavioral disorders. *Focus on Exceptional Children, 26*(2), 1–10.

Stage, S. A. (2007, March). Stage Observation System. Paper presented at the Annual National Association for School Psychologists Convention, New York, NY.

Sugai, G., Horner, R., & Gresham, S. (2002). Behaviorally effective school environments. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches.* Washington, DC: National Association of School Psychologists.

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

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

Wallace, J. M., Yamaguchi, R., Bachman, J. G., O'Malley, P. M., Schulenbert, J. E., & Johnston, L. D. (2007). Religiosity and adolescent substance use: The role of individual and contextual influences. *Social Problems, 54*(2), 308–327.

Wheaton, B., & Clarke, P. (2003). Space meets time: Integrating temporal and contextual influences on mental health in early adulthood. *American Sociological Review, 68*(5), 680–706.

Wiley, A., Siperstein, G., Forness, S., & Brigham, F. (2010). School context and the problem behavior and social skills of students with emotional disturbance. *Journal of Child and Family Studies, 19*(4), 451–461.

Zins, J. E., & Ponte, C. R. (1990). Best practices in school-based consultation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology—II* (pp. 673–694). Washington, DC: National Association of School Psychologists.

AUTHORS' NOTE

This research was supported by a grant from the US Department of Education, Institute of Education Sciences (R324A07183).

MANUSCRIPT

Initial Acceptance: 11/13/12
Final Acceptance: 12/28/13

APPENDIX

TABLE A1
Bivariate (Unadjusted) Intercorrelations among Baseline and Outcome Behavior

Measure	1	2	3	4	5	6	7	8	9
School baseline									
1. TRS externalizing	—	**.84**	**−.93**	.00	.00	.00	−.23	.10	.26
2. # problem behavior	**−.72**	—	**−.81**	.00	.00	.00	−.22	−.02	.22
3. % on-task behavior	−.02	.14	—	.00	.00	.00	**.37**	−.02	−.25
Student relative baseline									
4. TRS externalizing	.00	.00	.00	—	−.16	.11	**.42**	**.45**	−.12
5. # problem behavior	.00	.00	.00	.07	—	**−.71**	.08	.04	−.05
6. % on-task behavior	.00	.00	.00	.18	**−.47**	—	−.24	−.07	.10
Student final outcome									
7. TRS externalizing	.30	−.27	−.08	.38	.19	.27	—	**.43**	−.20
8. # problem behavior	−.16	.22	−.15	.11	−.05	.24	.18	—	**−.57**
9. % on-task behavior	−.02	.21	**.57**	−.07	.14	−.03	−.21	**−.69**	—

Note. N = 54 students (n = 35 treatment students in upper diagonal, n = 19 control students in lower diagonal). School Baseline = z-score of school aggregate baseline score; Student Relative Baseline = student z-score within school. Significant correlations ($p < .05$) are indicated in boldface.