Developmental Psychology

Prosocial Skills Causally Mediate the Relation Between Effective Classroom Management and Academic Competence: An Application of Direction Dependence Analysis

Wolfgang Wiedermann, Wendy M. Reinke, and Keith C. Herman

Online First Publication, July 23, 2020. http://dx.doi.org/10.1037/dev0001087

AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

http://dx.doi.org/10.1037/dev0001087

Prosocial Skills Causally Mediate the Relation Between Effective Classroom Management and Academic Competence: An Application of Direction Dependence Analysis

Wolfgang Wiedermann, Wendy M. Reinke, and Keith C. Herman
University of Missouri

Direction dependence analysis (DDA) is a recently developed method that addresses the need for more sophisticated tools to evaluate causal mechanisms of developmental processes and interventions. The present study applied DDA to evaluate the hypothesized mediators of a classroom behavior management training program on student academic competence. The study involved a group randomized controlled trial with 105 teachers and 1,818 students (K−3rd grade) in a large urban school district in the United States. Analyses revealed only student prosocial skill development causally mediated the intervention's effects on student academic competence. The findings support the importance of explicit instruction and coaching of student social skills as part of classroom behavior management programs and confirm the causal link between prosocial skills and academic success. The findings are discussed with regard to implications for future applications of DDA in developmental research.

Keywords: classroom management, teacher training, prosocial skills, academic competence, mediation, direction dependence analysis

Supplemental materials: http://dx.doi.org/10.1037/dev0001087.supp

Specifying causal relations is of central importance to developmental psychologists. However, as many authors have noted, causal inferences (i.e., moving from association to causation) are problematic in child development research given that most studies rely on cross-sectional (Holmbeck, Franks Bruno, & Jandasek, 2006) or naturalistic longitudinal designs (Foster, 2010; Larzelere, Gunnoe, Ferguson, & Roberts, 2019). Moreover, even when experimental studies are conducted, study designs often limit confidence in the proposed mechanisms of change (e.g., when proposed mediators are collected simultaneous to observed outcomes). Thus, the need for better tools for causal inference has been recognized by developmental psychologists. For example, Foster (2010) provided a critical review of the necessary assumptions of standard regression models to interpret results as being causal and introduced modern causal inference methods for observational data (such as propensity score and instrumental variable approaches) to the audience of developmental psychologists. The present article continues the discussion of statistical methods for causal inference in developmental research and uses a novel statistical framework, direction dependence analysis (DDA; Wiedermann & Li, 2018; Wiedermann & von Eye, 2015), to test causal mechanisms linking teacher classroom management skills to improvements in youth academic achievement.

Direction Dependence Analysis: Rationale and Overview

A critical component of establishing causal inferences in developmental sciences is the a priori specification and subsequent evaluation of a theory of change linking the developmental process or intervention to the outcome of interest via intermediary mediational processes (Gottfredson et al., 2015). The mediating factors in these theories are the proposed proximal mechanisms the intervention impacts that in turn produce the more distal youth outcomes.

Knowledge development paradigms emphasizing the confirmation of hypothesized mediating processes are limited by existing research designs and methodology (Gottfredson et al., 2015). In (cross-sectional) observational studies, for example, finding a statistically significant indirect effect does not prove the existence of causation. The reason for this is that (a) reversing the causal direction of paths in mediation models may be theoretically justified and standard regression approaches cannot be used to distinguish between causally competing mediation models (Wiedermann & von Eye, 2015, 2016) and (b) unobserved confounders may bias indirect effect estimates (Bullock, Green, & Ha, 2010). Experimental control and temporal precedence (the issue of using

Wolfgang Wiedermann, Wendy M. Reinke, and Keith C. Herman, Missouri Prevention Science Institute and Department of Educational, School, and Counseling Psychology, University of Missouri.

The research reported here was supported by the Institute of Educational Sciences, U.S. Department of Education, through Grant R305A100342 to Keith C. Herman and Wendy M. Reinke. The opinions expressed are those of the authors and do not represent views of the institute or the U.S. Department of Education.

Correspondence concerning this article should be addressed to Wolfgang Wiedermann, Missouri Prevention Science Institute and Department of Educational, School, and Counseling Psychology, University of Missouri, Hill Hall 13b, Columbia, MO 65211. E-mail: wiedermannw@missouri.edu

temporality to establish causal statements is taken up in the "Discussion" section) are commonly recommended to strengthen causal claims. However, even in the most rigorous research design for establishing causal relations, randomized controlled trials, only one variable is typically manipulated, exposure to the intervention.

Randomization rarely occurs at the mediator level. While blockage and enhancement manipulations of mediators have been proposed (Imai, Tingley, & Yamamoto, 2013; Pirlott & MacKinnon, 2016), these *manipulation-of-mediator designs* impose their own challenges and limitations (Bullock et al., 2010). For example, even when both the independent variable and the mediator are experimentally controlled, one cannot automatically rule out alternative explanations of the mediator-outcome relation (i.e., confounding mediating variables may still exist), and it can be challenging to manipulate the mediator at a precise level necessary to demonstrate mediation. Additionally, mediation models are often complex, with multiple intervening processes that may evolve over time, sometimes decades, necessitating the collection of many waves of data, which is often economically infeasible (Gottfredson et al., 2015). In reality, mediators are often collected at the same time as the outcomes (so-called *measurement-of-mediation designs*; Pirlott & MacKinnon, 2016), and mediational mechanisms are tested through performing statistical mediation analyses. Thus, causal inference of mediational analyses often suffers from the same limitations as cross-sectional correlational studies (Imai, Keele, & Tingley, 2010). The standard measurement-of-mediation design cannot differentiate whether (a) the mediator (m) causes the outcome (y; i.e., $m \rightarrow y$), (b) the outcome causes the mediator ($y \rightarrow m$), or (c) an unmeasured confounder (u) causes both variables (i.e., $m \leftarrow u \rightarrow y$; this includes both full and partial confounding).

DDA provides a promising tool for overcoming some of the enduring limitations of traditional mediation approaches (Wiedermann, Li, & von Eye, 2019; Wiedermann & Sebastian, 2019a; Wiedermann & von Eye, 2016). Compared to standard causal inference methods (e.g., propensity score and instrumental variable approaches), DDA differs in two important aspects: First, DDA is not primarily concerned with eliminating biases in causal effect estimates. Instead, DDA critically evaluates a causal mechanism through probing the model-implied causal direction of effects. Second, while traditional approaches to mediation usually assume that variables follow a normal (Gaussian) distribution and utilize variable information up to second-order moments (variances and covariances), DDA requires nonnormality of variables and makes use of third- and fourth-moment information (skewness, coskewness, kurtosis, and cokurtosis). The reason for this is that, under nonnormality of variables, a causal mechanism changes not only the means, variances, and covariances but also the distributional shape of the variables. These distributional changes can be used to distinguish competing explanatory models ($m \rightarrow y$, $y \rightarrow m$, or $m \leftarrow u \rightarrow y$).

DDA focuses on three different statistical components that are used for model selection: (a) distributional characteristics of observed variables, (b) distributional characteristics of residual terms of causally competing models, and (c) independence properties of independent variables and residuals in causally competing models. Combining the three components leads to a statistical framework that enables one to uniquely identify each one of the three explanatory models discussed above. Figure 1 summarizes DDA component patterns for causally competing models ($m \rightarrow y$ vs. $y \rightarrow m$) and a model where an unobserved confounder is present ($m \leftarrow u \rightarrow y$). In essence, in a correctly specified causal model, (a) the outcome will be closer to normality than the mediator, (b) the error (i.e., estimated model residuals) will be closer to normality than the residuals of the causally misspecified model, and (c) the independence assumption of mediators and errors (necessary to endow mediational effect estimates with causal meaning) will hold, while nonindependence will be observed in the misspecified model.

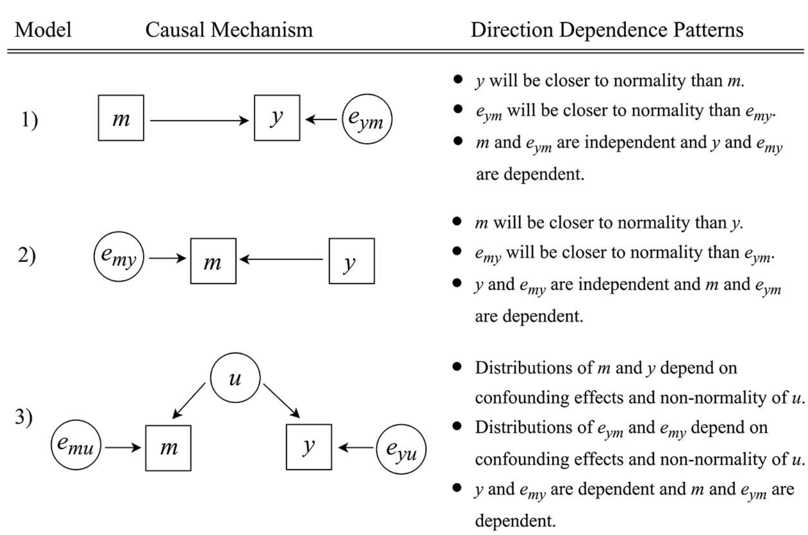

Figure 1. Causal mechanisms and corresponding direction dependence patterns for three possible explanations for an observed association of a mediator m and an outcome y (rectangles are used for observed variables and circles are used for unobserved variables; e = error term). Model 1: m is the cause and y is the effect. Model 2: y is the cause and m is the effect. Model 3: An unobserved confounder u is responsible for the association between m and y.

DDA Applied to Causal Mechanisms of Developmental Interventions

Most work to date using DDA has focused on establishing causal mechanisms of pairs of variables either because of the nature of collected data (cross-sectional) or because randomization to the causal process would be unethical (e.g., to determine if lead exposure causes attention-deficit/hyperactivity disorder). DDA may also be useful for evaluating proposed mediators of interventions intended to promote youth developmental outcomes. One application of particular relevance to the need for more sophisticated mediation tools concerns the relations between teacher classroom behavior management and student academic achievement. It is important to support teacher use of effective classroom management skills in part because student disruptive behaviors represent one of the most significant areas of stress and concerns voiced by teachers (Reinke, Stormont, Herman, Puri, & Goel, 2011). Effective classroom behavior management may not only reduce problem behavior and promote positive social and emotional development but also contribute to student academic performance. For instance, developmentally informed theories suggest that positive and predictable interactions with adults, such as teachers, are associated with emotion regulation and development of social competence, which in turn support academic achievement through more on-task behaviors and access to academic instruction (Patterson, Reid, & Dishion, 1992; Pianta, 1999).

Empirical studies support these theorized pathways. First, both cross-sectional and longitudinal studies support the hypothesized links between effective classroom management, positive classroom climates, and student achievement (Back, Polk, Keys, & McMahon, 2016; Reyes, Brackett, Rivers, White, & Salovey, 2012). For instance, Back and colleagues (2016) found that effective classroom management predicted increases in student performance on standardized achievement tests a year later. Likewise, many prior studies have shown that prosocial behaviors (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Gerbino et al., 2018; Malecki & Elliot, 2002; Wentzel, 1993), self-regulation (Durlak et al., 2011; Edossa, Schroeders, Weinert, & Artelt, 2018; Sawyer et al., 2015), concentration/attention problems (Rabiner, Godwin, & Dodge, 2016; Sayal, Washbrook, & Propper, 2015), and disruptive behaviors (Kremer, Flower, Huang, & Vaughn, 2016; Sayal et al., 2015) are predictive of concurrent and prospective academic performance. Second, prior experimental studies have demonstrated that classroom behavior management training programs lead to improvements in both student social development and academic outcomes (see Embry, 2002 and Kellam et al., 2011 for reviews). For instance, a recent group randomized trial in 102 classrooms with 1,450 students found that a classroom management program led to significant reductions in disruptive behaviors and concentration problems and also improvements in student standardized achievement scores in math and reading (Herman, Reinke, & Dong, 2018). Further, McCormick, Cappella, O'Connor, and McClowry (2015) found that improvements in classroom climate and organization predicted academic gains among kindergarten and first-grade students randomly assigned to receive a classroom-based social emotional intervention called INSIGHTS.

Incredible Years Teacher Classroom Management Training

One particularly promising classroom behavior management training program that targets students' social, emotional, and academic development is the Incredible Years Teacher Classroom Management (IY TCM) program. IY TCM was designed to promote the knowledge and use of effective classroom management practices among preschool and early elementary teachers, including effective praise, proactive teaching strategies (e.g., clear expectations, schedules, precorrection), giving effective commands, consistent consequences, planned ignoring, and use of time-out procedures. Additionally, the program focuses on ways to foster prosocial skills, positive relationships between teachers and parents, and increased parental involvement in their children's education and collaboration with teachers (Webster-Stratton, Reid, & Hammond, 2004).

The IY TCM program has been rated as "promising" by Blueprints for Healthy Youth Development (n.d.) and as "very promising" by the Office of Juvenile Justice and Delinquency Prevention's Model Programs (2018) based on the findings from several experimental studies. IY TCM has been evaluated in several randomized trials as part of a larger treatment package (Webster-Stratton, Reid, & Stoolmiller, 2008; Webster-Stratton, Reid, & Hammond, 2001; Webster-Stratton et al., 2004). These trials provided consistent evidence that IY TCM classrooms were more positive and nurturing and that students in these classrooms were less aggressive, more prosocial, and more on task compared to business-as-usual classrooms. More recently, several studies have evaluated the impact of the IY TCM program as a stand-alone intervention. Randomized controlled trials (RCTs) have found IY TCM causes improvements in student prosocial skills (Baker-Henningham, Scott, Jones, & Walker, 2012; Hutchings, Martin-Forbes, Daley, & Williams, 2013; Reinke, Herman, & Dong, 2018), emotional regulation (Reinke et al., 2018), concentration problems/off-task behaviors (Hutchings et al., 2013), and disruptive behaviors (Baker-Henningham et al., 2012). A recent cluster RCT called STARS found that students in IY TCM classrooms experienced improvements in overall teacher-rated behavior difficulties; however, these improvements were not sustained at longer-term follow-ups (Ford et al., 2019). Moreover, the STARS trial did not find significant improvements in teacher self-efficacy ratings or burnout (Hayes et al., 2020).

IY TCM's Theory of Change

Patterson et al.'s (1992) dual-failure (academic and social) model provided the theoretical foundation of IY TCM and for understanding the impact of the intervention on student behaviors and academic achievement. This model is one of the most studied and supported theories explaining the emergence of conduct problems and academic failure, emphasizing the importance of the family and teacher socialization processes, especially during the early years of development. Parents and teachers of children displaying early signs of aggression and challenging behaviors may fall into a cycle of negative reinforcement, which acts to maintain challenging behaviors through a coercive cycle in which parents and teachers acquiesce to child escalating demands. In turn, parents and teachers use harsh discipline practices when the child misbehaves, which in turn escalates to severe misbehavior.

IY TCM attempts to alter these social interaction patterns in early elementary school to produce positive student outcomes. In particular, IY TCM trains teachers to use more proactive strategies to interrupt coercive classroom processes. By increasing positive interactions with all students, IY TCM reduces disruptive behavior. Simultaneously, teachers explicitly model and teach social-emotional skills including prosocial behaviors and emotional regulation. Collectively, these strategies may increase student engagement, time on task, and time for instruction. With more exposure to instruction and opportunities for learning, IY TCM helps to increase student academic competence.

The IY TCM theory of change that guided program development and evaluation specifies student social and emotional behaviors as proximal outcomes of the intervention (see Webster-Stratton, 2018). In turn, the IY TCM theory of change predicts that these immediate intervention effects contribute to more distal outcomes including improvements in student academic competence. Much less empirical evidence has examined these distal outcomes of IY TCM. Of note, a recent RCT in 45 rural elementary classrooms reported no significant improvements in academic skills of students in IY TCM relative to comparison group students (Murray, Rabiner, Kuhn, Pan, & Sabet, 2018). However, the study also reported null effects on nearly all teacher and student outcomes, prompting the authors to conclude that high baseline classroom functioning may have left little room for improvement on any of the targeted variables in this sample.

Even more germane to the present study, we could find no prior study that has examined mediators of IY TCM outcomes. To our knowledge, all previous evaluations of IY TCM have only reported main or moderated effects. Thus, a critical void in the IY TCM literature is the lack of studies that have empirically evaluated the intervention's theory of change, particularly whether improvements in student social and emotional behaviors mediate program effects on academic achievement.

Current Study

The current study applied DDA to evaluate mechanisms that explain the effects of teacher classroom behavior management training on student outcomes. Using data from a prior study (Reinke, Herman, & Dong, 2018), we examined four student behavior outcomes as plausible mediators of this effect: prosocial behavior, self-regulation, concentration problems, and disruptive behaviors. In line with IY TCM's theory of change, we examined whether changes in proximal student behaviors mediated the effect of a classroom management program on student academic competence. We hypothesized that prosocial skills, self-regulation, concentration problems, and disruptive behaviors would each independently mediate IY TCM effects on student academic competence. In addition, we evaluated joint indirect effects of multiple mediation models to account for associations between mediators. Because posttreatment mediators and academic competence were measured at the same measurement occasion, reverse causation phenomena cannot be ruled out. For example, a causal mechanism in which IY TCM increases academic competence, which in turn increases prosocial skills, may also be plausible. Also, unconsidered confounders can bias both the causal direction and the magnitude of mediation effects. Thus, we used DDA to evaluate the plausibility of causally competing mediation models and to test the absence/presence of influential confounders.

Method

Participants and Procedure

Data for the current analyses were drawn from a study that assessed the impact of the IY TCM program on student social-emotional skills, disruptive behavior, and academic outcomes using a large-scale RCT (Reinke et al., 2018). Participants in the study were 105 teachers and 1,817 students in kindergarten to third grade from nine schools in a school district in the midwestern part of the United States. This study was approved by the institutional review board at the University of Missouri (project title: "Evaluation of a video-based modeling program to promote effective teacher classroom management practices," Protocol 1163417C). A block cluster randomized design was used to randomly assign teachers to receive IY TCM or business as usual. The majority of students in the study qualified for free or reduced lunch (61%) and identified as Black (76%; 22% White, 2% other). The sample included slightly more male students (52%), and about 9% received special education service. The analysis sample included $n = 1{,}648$ students (90%; 104 teachers) who had valid baseline and posttreatment measures. No significant demographic differences were found between the analysis sample and the subsample with incomplete data. Table 1 shows the descriptive statistics for student- and classroom-level variables for the overall sample and by treatment status.

The study had high rates of enrollment (96% of eligible teachers and 85% of eligible students provided enrolled) and retention (100% of teachers and 93% of students provided follow-up data). Teachers assigned to the IY TCM condition attended 6 full days of training provided over a 3-month period. The trainings were led by doctoral-level trainers supervised by the program developer. A doctoral-level IY TCM coach supported teacher implementation outside of the training sessions. The coach visited each teacher seven times, on average; these visits included classroom observations and providing individual support and feedback. Fidelity of implementation was demonstrated by high rates of teacher attendance at the trainings (training attendance ranged from 94–100%) and time spent with the coach (6 hr per teacher, on average). Additionally, independent observers conducted observations in each classroom three times after the initial training session and documented significant increases in teacher use of proactive teacher strategies versus teachers in the comparison condition (see Reinke et al., 2018 for full report on training protocol and implementation fidelity).

Student-Level Measures

Student behavior and academic competence. The Teacher Observation of Classroom Adaptation-Checklist (TOCA-C; Koth, Bradshaw, & Leaf, 2009)—a 54-item measure completed by teachers to rate child behavior in the past 3 weeks—was used to assess concentration problems (e.g., "easily distracted" and "stays on task"), disruptive behavior (e.g., "breaks rules" and "fights"), prosocial behavior (e.g., "shows empathy" and "is friendly"), and

Table 1
Descriptives for Student- and Classroom-Level Sample

Level of analysis	Treatment (n = 822 students in 52 classrooms)	Control (n = 821 students in 52 classrooms)	Overall (n = 1,643 students in 104 classrooms)
Student level			
Age *M* (*SD*)	7.15 (1.22)	7.06 (1.1)	7.1 (1.16)
Female *n* (%)	392 (47.8)	404 (49.2)	796 (48.4)
Black *n* (%)	626 (76.2)	613 (74.7)	1239 (75.4)
Reduced lunch *n* (%)	493 (60.0)	501 (61.0)	994 (60.5)
Special education *n* (%)	70 (8.5)	80 (9.7)	150 (9.1)
Academic competence (pre) *M* (*SD*)	3.22 (1.28)	3.22 (1.25)	3.22 (1.27)
Concentration problems (pre) *M* (*SD*)	3.13 (1.32)	3.18 (1.29)	3.15 (1.31)
Disruptive behavior (pre) *M* (*SD*)	1.75 (0.78)	1.78 (0.74)	1.76 (0.76)
Emotional dysregulation (pre) *M* (*SD*)	2.28 (1.01)	2.29 (0.95)	2.29 (0.98)
Prosocial behavior (pre) *M* (*SD*)	4.52 (0.98)	4.44 (0.98)	4.48 (0.98)
Academic competence (post) *M* (*SD*)	3.53 (1.25)	3.44 (1.28)	3.48 (1.27)
Concentration problems (post) *M* (*SD*)	2.6 (1.24)	2.67 (1.24)	2.64 (1.24)
Disruptive behavior (post) *M* (*SD*)	1.85 (0.83)	1.90 (0.79)	1.88 (0.81)
Emotional dysregulation (post) *M* (*SD*)	2.15 (1.08)	2.29 (1.08)	2.22 (1.08)
Prosocial behavior (post) *M* (*SD*)	4.84 (1.04)	4.67 (1.06)	4.75 (1.05)
Classroom level			
Average age *M* (*SD*)	7.16 (1.16)	7.09 (1.05)	7.12 (1.10)
Number of students *M* (*SD*)	17.31 (3.28)	17.29 (3.26)	17.30 (3.25)
Grade *n* (%)			
K	14 (13.5)	14 (13.5)	28 (26.9)
First	15 (14.4)	14 (13.5)	29 (27.9)
Second	10 (9.6)	16 (15.4)	26 (25.0)
Third	13 (12.5)	8 (7.7)	21 (20.2)
Percent female *M* (*SD*)	47.67 (8.41)	49.23 (9.43)	48.45 (8.92)
Percent Black *M* (*SD*)	76.29 (25.87)	75.92 (25.88)	76.10 (25.75)
Percent reduced lunch *M* (*SD*)	61.22 (24.78)	62.97 (23.70)	62.10 (24.14)
Percent special education *M* (*SD*)	8.62 (8.20)	9.64 (7.57)	9.13 (7.87)
Average academic competence (pre) *M* (*SD*)	3.19 (0.58)	3.19 (0.51)	3.19 (0.54)
Average concentration problems (pre) *M* (*SD*)	3.14 (0.79)	3.16 (0.78)	3.15 (0.78)
Average disruptive behavior (pre) *M* (*SD*)	1.77 (0.35)	1.79 (0.33)	1.78 (0.34)
Average emotional dysregulation (pre) *M* (*SD*)	2.30 (0.55)	2.30 (0.47)	2.30 (0.51)
Average prosocial behavior (pre) *M* (*SD*)	4.50 (0.59)	4.42 (0.59)	4.46 (0.59)
School *n* (%)			
I	5 (4.8)	6 (5.8)	11 (10.6)
II	7 (6.7)	6 (5.8)	13 (12.5)
III	6 (5.8)	6 (5.8)	12 (11.5)
IV	5 (4.8)	6 (5.8)	11 (10.6)
V	6 (5.8)	5 (4.8)	11 (10.6)
VI	6 (5.8)	6 (5.8)	12 (11.5)
VII	5 (4.8)	6 (5.8)	11 (10.6)
VIII	7 (6.7)	6 (5.8)	13 (12.5)
IX	5 (4.8)	5 (4.8)	10 (9.6)

emotional dysregulation (e.g., "easily frustrated" and "changes moods quickly"). Item responses ranged from *never* (1) to *almost always* (6). Mean scores based on subscale-specific items were used to measure child behavior. Cronbach's alpha for each subscale ranged from .77 to .96. Student behavior variables significantly deviated from normality (all Shapiro-Wilk *p*s $< .001$), with skewness ranging from -0.59 to 1.19 and excess-kurtosis values ranging from -0.91 to 1.27.

Students' academic competence was measured using the corresponding subscale of the Revised Teacher Social Competence Scale (Gifford-Smith, 2000). The academic competence subscale consists of seven items (e.g., "Able to read grade level material" or "Performing academically at grade level"), with responses ranging from *almost never* (0) to *almost always* (5; Cronbach's alpha = .93). Academic competence scores significantly deviated from normality (Shapiro-Wilk $p < .001$), with a skewness of -0.53 and an excess-kurtosis value of -0.62.

Because student behavior and academic outcome measures were based on teacher reports, we evaluated the potential presence of a common-rater bias. Following Podsakoff, MacKenzie, Lee, and Podsakoff (2003), we conducted a factor analysis on all the student behavior and academic outcome construct items (known as Harman's one-factor test) to evaluate the magnitude of a common-rater bias. The one-factor model accounted for about 49% of the variance. In addition, several fit indices (such as Velicer's minimum average partial test, the Bayes information criterion, and the root mean square error of approximation) clearly rejected the one-factor model, indicating that the common-rater bias might not be substantial in the present data.

Student-level covariates. Sex (0 = male, 1 = female), race (0 = White/Other, 1 = Black), special education service (0 = no, 1 = yes), free/reduced lunch status (0 = no, 1 =yes), and age (in yrs.) were used as student-level covariates. Data were obtained from the school district for all participating students.

Classroom-Level Measures and Covariates

Aggregated Level-1 variables were included as Level-2 predictors to study potential contextual effects (Raudenbush, 1989) and to reduce potential biases in Level-1 estimates (Huang, 2018). Specifically, mean aggregated scores for the TOCA-C subscales and academic competence and aggregated information on student demographics (average age, percent Black students, percent female students, percent of students qualifying for free/reduced lunch, percent of students with special education service, and the number of students per classroom) were included. Grade-level information was also included as a covariate (reference = kindergarten). School affiliation dummy variables were used to account for nesting of teachers within schools and to handle all possible (measured and unmeasured) school-level predictors (cf. McNeish & Kelley, 2019).

Overview of Statistical Models and Method

Mediation analysis. To examine mediational paths between the IY TCM treatment effects and academic competence, we conducted a multilevel mediation analysis using 2–1–1 models (Pituch & Stapleton, 2012) in which students' posttreatment behavior (Level 1) mediated the association between classroom intervention (Level 2) and posttreatment academic competence (Level 1). The equations for the mediator and outcome models are

Mediator model:

$$m_{ij} = \beta_{0j} + \sum \beta_{kj} x_{ijk} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01} T_j + \sum \gamma_{0g} w_g + u_{0j}$$

Outcome model:

$$y_{ij} = \beta'_{0j} + \beta_1 m_{ij} + \sum \beta'_{kj} x_{ijk} + r'_{ij}$$
$$\beta'_{0j} = \gamma'_{00} + \gamma'_{01} T_j + \gamma_{02} \overline{m}_j + \sum \gamma'_{0g} w_g + u_{0j}$$

with m_{ij} being the mediator (i.e., the posttreatment TOCA-C composite scores) of student i in class j, x_{ijk} collecting Level-1 baseline measures (pretreatment TOCA-C and academic competence measures) and student-level covariates, T_j being the treatment status of class j (0 = control, 1 = IY TCM program), and w_g defining Level-2 covariates (i.e., average pretreatment TOCA-C and academic competence measures and aggregated demographic information). The product coefficient $\gamma_{01}\beta_1$ is an estimate for the cross-level indirect effect, $\gamma_{01}\gamma_{02}$ estimates the contextual (cluster-level) indirect effect, and $\gamma_{01}\beta_1 + \gamma_{01}\gamma_{02}$ represents the total indirect effect (Krull & MacKinnon, 2001; Pituch & Stapleton, 2012). Cluster bootstrapping (cf. Huang, 2016; with 2,000 resamples) was used to compute 95% percentile bootstrap confidence intervals to evaluate the significance of indirect effects. Likelihood ratio χ^2 tests were used to evaluate whether random effects for the mediators were necessary. Potential treatment-mediator interactions were tested using the outcome model

$$y_{ij} = \beta'_{0j} + \beta_{1j} m_{ij} + \sum \beta'_{kj} x_{ijk} + r'_{ij}$$
$$\beta'_{0j} = \gamma'_{00} + \gamma'_{01} T_j + \gamma_{02} \overline{m}_j + \sum \gamma'_{0g} w_g + u_{0j}$$
$$\beta_{1j} = \gamma'_{10} + \gamma'_{11} T_j + u_{1j}$$

with γ_{11}' being an estimate for the treatment-mediator interaction. All mediation analyses were performed using R 3.4.3 (R Core Team, 2019) and the lme4 package (Version 1.1–15; Bates, Mächler, Bolker, & Walker, 2015).

Direction dependence analysis. In the methodological development of direction dependence modeling techniques, previous studies predominantly focused on single-level data (e.g., Li & Wiedermann, 2019; von Eye & DeShon, 2012). DDA has not yet been proposed for nested (multilevel) data. To apply DDA in the context of 2–1–1 multilevel mediation models, we made use of the fact that the mediator and the outcome are both measured at the student level. Therefore, after properly accounting for teacher-level effects, standard DDA can be applied. We used a fixed-effects approach to account for teacher-level effects. That is, to evaluate direction dependence properties of the mediator-outcome paths, mediators and outcome were regressed on all Level-1 covariates and $h - 1$ fixed effects to account for (Level-2) variation of the h classrooms,

$$m_{ij} = \beta_0 + \sum \beta_{jk} x_{ijk} + \sum \beta_j (Teacher_j) + r_{ij}^{(m)}$$
$$y_{ij} = \beta_0 + \sum \beta_{jk} x_{ijk} + \sum \beta_j (Teacher_j) + r_{ij}^{(y)}.$$

The extracted residuals $r_{ij}^{(m)}$ and $r_{ij}^{(y)}$ of the two auxiliary models represent covariate-adjusted mediator and outcome measures, which were subsequently used to evaluate causal effect directionality (note that both teacher- and school-level effects are completely accounted for through including teacher fixed effects). Normality tests (i.e., D'Agostino skewness and Anscombe-Glynn kurtosis tests) and 95% bootstrap confidence intervals (with 2,000 resamples) for differences in skewness and kurtosis values were used to evaluate distributional characteristics of $r_{ij}^{(m)}$ and $r_{ij}^{(y)}$ and estimated residuals of the two competing models $r_{ij}^{(m)} \rightarrow r_{ij}^{(y)}$ and $r_{ij}^{(y)} \rightarrow r_{ij}^{(m)}$. Similarly, 95% bootstrap confidence intervals (with 2,000 resamples) were used to evaluate differences in coskewnesses of $r_{ij}^{(m)}$ and $r_{ij}^{(y)}$ (i.e., measures of asymmetry of distributions that take into account the relatedness of y and m; cf. Wiedermann et al., 2019; Wiedermann & Sebastian, 2019b). The Breusch-Pagan homoscedasticity test (BP test; Breusch & Pagan, 1979) and the Hilbert-Schmidt independence criterion (HSIC; an omnibus measure of independence; cf. Gretton et al., 2008) were applied to evaluate independence properties of competing models. DDA model selection was performed using the decision guidelines presented in Figure 1. Specifically, the hypothesized mediation mechanism IY TCM → student behavior → academic outcome finds empirical support when (a) the distribution of (covariate-adjusted) academic competence scores is closer to normality than the distributions of (covariate-adjusted) student behavior variables, (b) the residuals of the model that treats academic competence as the dependent variable are closer to the normal distribution than the residuals of a model that treats academic competence as an independent variable, and (c) the independence assumption holds in the model that uses academic competence as the dependent variable and, at the same time, a violation of the independence assumption is observed for the causally reversed models. Simulation-based sensitivity analysis approaches described in Wiedermann, Merkle, and von Eye (2018) and Wiedermann and Sebastian (2019b) were used to evaluate the robustness of DDA model selection against measurement error and additional (unobserved) confounding. DDA tests were performed using R (Wiedermann & Li, 2019; see www.ddaproject.com).

Results

Figure 2 gives the path diagrams of the four mediation models in which prosocial behavior, emotional dysregulation, concentration problems, and disruptive behavior were separately considered as mediators. Covariate-adjusted regression coefficients, standard errors, 95% confidence intervals, and p values of the mediator and outcome models are given in Table 2 (detailed results of the multilevel mediation models are given in the online supplemental materials). No indirect effects were observed for concentration problems (total indirect effect = 0.052, 95% CI [−0.032, 0.144]; see Figure 2a) and disruptive behavior (total indirect effect = 0.020, [−0.025, 0.059]; see Figure 2b). Significant indirect effects were observed for prosocial behavior (total indirect effect = 0.089, [0.009, 0.171]) and emotional dysregulation (total indirect effect = 0.090, [0.021, 0.166]). The total indirect effect of prosocial behavior is attributable to a significant cross-level indirect effect (indirect effect = 0.058, [0.006, 0.104]); that is, the IY TCM treatment increased individuals' prosocial behavior ($\beta = 0.14$, $SE = 0.06$, $p = .025$), which in turn increased academic competence ($\beta = 0.40$, $SE = 0.03$, $p < .001$). The corresponding contextual indirect effect did not reach significance (indirect effect = 0.027, [−0.005, 0.072]). No direct treatment effect was observed for academic competence ($\beta = 0.02$, $SE = 0.06$, $p = .773$), suggesting that prosocial behavior completely mediates the effect of the IY TCM treatment on academic competence. Similarly, the total indirect effect of emotional dysregulation is explained by a significant cross-level mediation mechanism (indirect effect = 0.043, [0.016, 0.067]) and a nonsignificant contextual indirect effect (indirect effect = 0.048, [−0.012, 0.112]). The IY TCM program significantly reduced individuals' emotional dysregulation ($\beta = -0.18$, $SE = 0.05$, $p = .001$), which in turn increased academic competence ($\beta = -0.24$, $SE = 0.02$, $p < .001$). Again, individuals' emotional dysregulation completely mediated the effect of the IY TCM treatment on academic competence (direct effect: $\beta = 0.01$, $SE = 0.07$, $p = .840$).

When both mediators were considered simultaneously in a multiple mediation model, the indirect effect of emotional dysregulation was no longer significant (total indirect effect = 0.021, 95% CI [−0.032, 0.074]), while the indirect effect of prosocial behavior remained significant (total indirect effect = 0.081, [0.012, 0.164]). Further, when all four mediators were considered simultaneously in a multiple mediation analysis, again, only the indirect effect of prosocial behavior was significant (total indirect effect = 0.052, [0.002, 0.117]). Detailed results of the full multiple mediation model are given in the online supplemental materials. This suggests that the IY TCM program increases academic competence through increasing prosocial behavior even after controlling for mediational influences of the other student behavior components. Treatment-mediator interactions were nonsignificant in all models (all $ps > .22$).

Next, we evaluated the causal direction of the relation between the mediators (prosocial behavior [PB] and emotional dysregulation [ED]) and academic competence using DDA. In the first step, measures of PB, ED, and academic competence (AC) were residualized using the regression models described above. That is, PB, ED, and AC were regressed on all student-level covariates and on teacher fixed effects, and extracted residuals (reflecting adjusted PB, ED, and AC variables) were subsequently used in DDA. In line with the results in Table 2, covariate-adjusted measures of AC and PB were positively related ($\beta = 0.41$, $SE = 0.03$, $p < .001$, $R^2 = 0.14$). Linearity of the effect was confirmed by inspecting LOWESS plots (Figure S1 in the online supplemental materials) and testing higher-order terms (adding a quadratic effect for PB led to a nonsignificant R^2 increase; $p > .50$). Because no distinct causal decisions were possible for kurtosis-based DDA measures, we focused on skewness-related information of variables. Detailed DDA results are given in Table 3. First, AC was symmetrically distributed ($p = .443$), while PB was significantly skewed ($p < .001$), which suggested a model of the form PB → AC. Similarly, both differences in skewness estimates and coskewnesses sug-

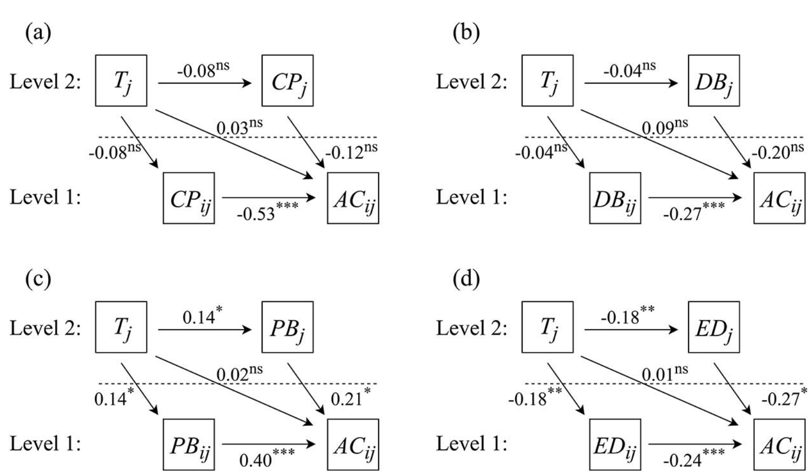

Figure 2. Multilevel mediation models for concentration problems (CP), disruptive behavior (DB), prosocial behavior (PB), and emotional dysregulation (ED) as mediators (T = treatment; AC = academic competence; i = student-level index; j = classroom-level index). ns = nonsignificant. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 2
Covariate-Adjusted Results for the Multilevel Mediation Models to Predict Academic Competence With Concentration Problems, Disruptive Behavior, Prosocial Behavior, and Emotional Dysregulation as Mediators

	Mediator model			Outcome model		
Variable	β (*SE*)	95% CI	*p*	β (*SE*)	95% CI	*p*
Concentration problems						
Treatment	−0.08 (0.07)	[−0.19, 0.04]	.229	0.03 (0.05)	[−0.06, 0.12]	.586
Concentration problems	—	—	—	−0.53 (0.02)	[−0.56, −0.49]	<.001
Average concentration problems	—	—	—	−0.12 (0.09)	[−0.27, 0.03]	.174
Disruptive behavior						
Treatment	−0.04 (0.05)	[−0.12, 0.04]	.365	0.09 (0.06)	[−0.03, 0.20]	.192
Disruptive behavior	—	—	—	−0.27 (0.03)	[−0.34, −0.21]	<.001
Average disruptive behavior	—	—	—	−0.20 (0.16)	[−0.47, 0.07]	.209
Prosocial behavior						
Treatment	0.14 (0.06)	[0.03, 0.25]	.025	0.02 (0.06)	[−0.08, 0.12]	.773
Prosocial behavior	—	—	—	0.40 (0.03)	[0.35, 0.45]	<.001
Average prosocial behavior	—	—	—	0.21 (0.10)	[0.04, 0.39]	.039
Emotional dysregulation						
Treatment	−0.18 (0.05)	[−0.27, −0.09]	.001	0.01 (0.07)	[−0.10, 0.13]	.840
Emotional dysregulation	—	—	—	−0.24 (0.02)	[−0.29, −0.19]	<.001
Average emotional dysregulation	—	—	—	−0.27 (0.13)	[−0.50, −0.04]	.048

Note. Unstandardized coefficients are reported. *SE* = standard error; CI = confidence interval. Based on n = 1,643 students in 104 classes (Level-1 covariates: gender, age, race, reduced lunch, special education, pretreatment academic competence, and pretreatment concentration problems, disruptive behavior, prosocial behavior, and emotional dysregulation; Level-2 covariates: grade, school, % female, % Black, % reduced lunch, % special education, average pretreatment academic competence, concentration problems, disruptive behavior, prosocial behavior, and emotional dysregulation).

gested that PB → AC is better suited to describe the causal flow than AC → PB. Results for distributional characteristics of model residuals pointed in the same direction, that is, residuals for PB → AC are symmetrically distributed (p = .240), while residuals of the reversed model are significantly skewed (p < .001). The difference in residual skewness was not significant. Next, we evaluated the independence assumption of the competing models to rule out the presence of influential confounders. Both the BP and the HSIC

Table 3
DDA Results for Covariate-Adjusted Student Behavior (Prosocial Behavior and Emotional Dysregulation) and Academic Competence

DDA component	Target model	Alternative model
	Variable relation: Prosocial behavior - academic competence	
Observed variable distributions		
Skewness of response	$\gamma = .05, z = 0.77, p = .443$	$\gamma = -.36, z = -5.78, p < .001$
Skewness differences[a]	$\Delta = 0.312$, 95% CI [0.146, 0.525]	
Higher-order correlation[a]	$\Delta = 0.021$, 95% CI [0.002, 0.057]	
Residual distributions		
Skewness	$\gamma = .07, z = 1.17, p = .240$	$\gamma = -.26, z = -4.30, p < .001$
Skewness differences[a]	$\Delta = 0.193$, 95% CI [−0.004, 0.437]	
Independence		
BP test	$\chi^2(1) = 0.15, p = .695$	$\chi^2(1) = 9.07, p = .003$
HSIC test	HSIC = .440, p = .168	HSIC = .894, p < .001
	Variable relation: Emotional dysregulation - academic competence	
Observed variable distributions		
Skewness of response	$\gamma = .05, z = 0.77, p = .444$	$\gamma = .77, z = 11.04, p < .001$
Skewness differences[a]	$\Delta = 0.693$, 95% CI [0.429, 1.030]	
Higher-order correlation[a]	$\Delta = 1.828$, 95% CI [0.866, 3.077]	
Residual distributions		
Skewness	$\gamma = .06, z = 0.98, p = .325$	$\gamma = .68, z = 10.27, p < .001$
Skewness differences[a]	$\Delta = 0.618$, 95% CI [0.349, 0.947]	
Independence		
BP test	$\chi^2(1) = 0.01, p = .950$	$\chi^2(1) = 9.70, p = .002$
HSIC test	HSIC = 0.837, p = .001	HSIC = 1.548, p < .001

Note. DDA = direction dependence analysis; BP test = Breusch-Pagan homoscedasticity test; HSIC = Hilbert-Schmidt independence criterion. Upper panel: target model = prosocial behavior → academic competence, alternative model = academic competence → prosocial behavior; lower panel: target model = emotional dysregulation → academic competence, alternative model = academic competence → emotional dysregulation. Based on n = 1,643 students in 104 classes.
[a] Values larger than zero suggest the corresponding target model.

test suggest that independence holds in the target model PB → AC and is violated in the reversed model AC → PB. Taken together, all DDA components suggest that PB causally affects AC and not vice versa. In the last step, we evaluated the robustness of DDA model selection against additional (unobserved) confounding and measurement error. Overall, the selected model PB → AC is quite robust against additional confounding; that is, about 20–25% of additional confounding is needed to render the two competing models indistinguishable. In addition, while accounting for measurement error in PB, a reliability of AC as low as 0.48 would be needed to conclude that distributional properties of variables no longer support the model PB → AC (details are given in the online supplemental materials).

To complete the analysis, we evaluated the causal direction of the association between ED and AC. As expected, both measures were negatively related ($\beta = -0.24$, $SE = 0.03$, $p < .001$, $R^2 = 0.06$), and LOWESS estimates (see Figure S2 in the online supplemental materials) and testing higher order terms suggested that the linearity assumption is satisfied ($p > .98$). The lower panel of Table 3 summarizes DDA results for residualized ED and AC scores. While DDA indicators for observed variable and residual distributions again preferred a causal model of the form ED → AC, the HSIC test rejected the independence assumption in both models. In other words, error dependence points at the presence of influential confounders. Because confounders can affect both the magnitude and the direction of causal effects, no clear-cut causal decision for ED and AC can be made for the present sample.

Discussion

Recently, authors have provided suggestions for using methods developed in other fields to improve confidence in causal inference in developmental research (e.g., Foster, 2010). The current article builds upon these suggestions by describing a statistical framework for addressing challenges with inferring the causal structure of variable relations. Using DDA, the present study confirmed one hypothesized causal mediator of IY TCM's effects on student academic competence.

The study findings advance knowledge about the IY TCM program. Most notably, this is the first study to our knowledge that has systematically tested the IY TCM theory of change. Consistent with the study hypothesis, student behavior mediated the effects of the classroom management training program on student gains in academic competence. However, a causal mediation process was only observed for one of the four hypothesized mediators, prosocial behavior. The IY TCM specifies the development of student prosocial skills as a proximal effect of the intervention and a precursor to improvements in student academic competence (Webster-Stratton, 2018). In addition to IY TCM's focus on providing structured and predictable environments, the program trains teachers to provide explicit social and emotional coaching and instruction to develop these skills. The present findings confirm this important aspect of the program and its downstream effects on student learning.

On the other hand, concentration problems and disruptive behavior were not supported as mediators by the present analysis. The null findings here were driven by the nonsignificant direct effects of treatment condition on these student outcomes. That is, although both concentration problems and disruptive behaviors had significant direct effects on the academic outcome, IY TCM did not improve student problem behaviors, and thus the indirect effect was not significant. One possible explanation for the lack of treatment effect on student problem behaviors has to do with the setting where the study occurred. Schools in this study were already implementing with high fidelity a School-Wide Positive Behavior Interventions and Supports (SW-PBIS) program as part of their routine practice. Prior studies have shown SW-PBIS significantly reduces student disruptive behaviors and concentration problems (Bradshaw, Waasdorp, & Leaf, 2012); thus, it is possible that this school-wide program lowered the overall rates of student disruptive behaviors and thus allowed teachers to prioritize their focus on student social skill development. Indeed, the baseline mean disruptive behavior scores for this sample were significantly lower than those reported in the TOCA-C development sample (Koth et al., 2009). Consistent with this explanation, a recent meta-analysis of IY TCM found improvements in disruptive behaviors only for youth with elevated levels of such behaviors at baseline (Nye, Melendez-Torres, & Gardner, 2018).

In contrast, emotional dysregulation was significantly associated with treatment status and students' academic competence. Thus, IY TCM had a direct effect on emotional dysregulation. As noted, IY TCM has a social-emotional coaching component whereby teachers support student coping skill development. These strategies may be key in reducing student emotional outbursts but may be less impactful in improving student concentration problems and other disruptive behaviors. Despite the direct effect of IY TCM, the unique causal role of emotional dysregulation as a mediator was not confirmed by DDA. The reason for this is that unconsidered influential confounders were likely to be present, as indicated by DDA. For instance, other academic enablers, such as student engagement and motivation, were not measured in this study and so could not be included in the analyses (Elliott, DiPerna, Mroch, & Lang, 2004).

Within the era of school accountability, the focus often has been on student achievement and teacher instructional skill, particularly in math, reading, and science. The present findings suggest that it is not only instructional skill and academic content that influence student academic competence. The social environment of the classroom, largely shaped by teacher behaviors, has a vital role in improving youth developmental outcomes. Here, creating classroom environments that support student prosocial competence was the variable that most prominently mediated the training effects on student academic competence in both simple and multiple mediation models.

The findings are largely consistent with the IY TCM theory of change and also abundant prior research showing concurrent and prospective links between social competence and academic achievement (Gerbino et al., 2018; Malecki & Elliot, 2002; Wentzel, 1993). However, the findings also imply the need for a more nuanced theory of change to describe the sequence of change in response to the program and to highlight how baseline characteristics of teachers and students implementing the program may influence outcomes. In classrooms with high levels of problem student behaviors, prior studies suggest that IY TCM may have proximal effects on reducing those behaviors and yet may only have limited to no effects on their prosocial behaviors (Nye et al., 2018). However, in settings such as the current study, where baseline levels of disruptive behaviors were relatively low, teachers may have more opportunities to coach and practice the social and emotional coping skills and in turn improve student prosocial behaviors and emotional regulation skills. In turn, the present study suggests that prosocial behaviors are the most proximal cause of student improvements in academic skills. One potential implication of this proposed sequence is that IY TCM may only impact student academic performance in contexts where disruptive behaviors are low; thus, implementing a school-wide behavior support program to reduce overall levels of disruptive behaviors and/or implementing IY TCM for a longer period of time (e.g., with perhaps a second year of training and support) in high-risk contexts may be needed to achieve improvements in student prosocial behaviors and academic outcomes.

One of the most important implications of the study is that the DDA method shows promise as a tool for examining other developmental processes and interventions. These may include examining whole school interventions such as multitiered academic and behavior support interventions. Entire social system interventions are particularly complex with many hypothesized mediators at multiple levels. The challenge of collecting multiple mediator data at repeated waves over time is magnified with these multitiered interventions. As one example, consider the attempts to evaluate the effects of a principal training program on student outcomes. A primary hypothesized mediator of these programs' effects on students is improvements in school climate and safety. Even getting to this first prominent mediator, however, requires other proximal mediators including principal uptake of the intervention; improvements in staff skill and morale in response to new principal behaviors; higher rates of staff involvement, monitoring, and interaction with students; and student experience of positive interactions and empowerment. In a traditional approach, each of these potential mechanisms would need to be measured at the exact time points sequenced in line with the theory of change. Using DDA, however, researchers could get a reasonable approximation of the sequence of events. This would allow researchers to feasibly test the complex steps from changes in leadership behavior and actions to youth social development and academic performance.

From a methodological perspective, the present study is, to our knowledge, the first to describe and apply direction dependence principles in the context of multilevel mediation models. Fixed effects were used to adjust student-level mediators and outcome for teacher-level effects. This approach has the advantage that it handles all measured and unmeasured Level-2 variables (McNeish & Kelley, 2019). Performing DDA for adjusted Level-1 variables enables one to detect potential confounding and to probe causal effect directionality while accounting for nested data structures. Although the present approach is restricted to 2–1–1 mediation processes, it provides a first important step to fully extend DDA to models for clustered data situations.

This study is not without some limitations. First, from a methodological perspective, DDA, of course, cannot and is not intended to replace well-established principles of experimentation. RCTs (deeply linked with the counterfactual/potential outcome framework of causation; Lewis, 1973) are arguably the gold standard to unpack causal mechanisms and are most credible for quantifying causal effects according to the standards of the What Works Clearinghouse (U.S. Department of Education, 2014). While no single theory or universal definition of "causality" exists (Beebee, Hitchcock, & Menzies, 2009), the evidence of causal mechanisms provided by methods of direction dependence is certainly weaker compared to the evidence obtained from RCTs. However, when RCTs are not feasible or difficult to implement, DDA provides a diagnostic toolkit to test competing causal explanations and identifies the best-fitting explanatory model based on the distributional properties of the data. That is, based on a set of candidate theories (derived using substantive knowledge and evidence), the one model that best explains the data is selected.

Second, regardless of the employed philosophical framework of causality and the statistical method of causal inference, causation can never be established by a single study alone. The present study provides the first evidence for a causal chain of the form teachers' behavioral skills → student prosocial behavior → student academic outcome. Replicability of the present results (e.g., replicating the study across different grades and various cultural contexts) and employing different causal inference techniques with nonoverlapping statistical assumptions are needed to strengthen the causal conclusions derived from the present sample. One implication is that "causal triangulation" (i.e., the process of strengthening causal claims by integrating results from various approaches, where each approach has different key sources of potential biases; Rosenström & García-Velázquez, 2020) is perhaps the most promising way to demonstrate robustness and validity of the proposed causal chain.

Third, findings from the study are based on measures of teacher report. Teachers were also the recipients of the training to implement IY TCM practices, leading to the possibility that teachers who received training may have rated their students as improved due to being exposed to the intervention. Further, we cannot rule out common rater effects in the relationship of mediators and outcome. Here, shared variance can either deflate or inflate the observed mediator-outcome relation (Podsakoff et al., 2003). However, Harman's one-factor test indicated that a common-rater bias might not be substantial in the present sample. Despite this, teachers are the most common source of information used to assess social behavior and determine special education evaluations (Zima et al., 2005); thus, their ratings are important in the context of school-based interventions and have been shown to predict social behavioral problems (Koth et al., 2009; Reinke, Herman, Petras, & Ialongo, 2008; Schaeffer, Petras, Ialongo, Poduska, & Kellam, 2003).

Fourth, the present analysis did not account for measurement error in composite measures of student behaviors and academic outcomes. Wiedermann, Merkle, and von Eye (2018) showed that measurement error can bias DDA decisions in terms of both the magnitude and the direction of the effect and proposed methods of moments (Fuller, 1987) and sensitivity analyses to account for measurement error in DDA measures that focus on observed variable distributions. Our results suggest that direction dependence decisions were highly robust against potential measurement error (detailed results are given in the online supplemental materials). Thus, we conclude that the impact of measurement error can be considered negligible in the present sample.

Finally, while the present study evaluated the causal nature of student behavior and academic competence using a measurement-of-mediator design (in which mediators and outcome were collected at the same time and indirect effect estimates were adjusted for baseline mediators and outcome measures), sequential mediation designs (i.e., the mediator is measured prior to the outcome) are often viewed as essential to establish statements of causal mediation (Kraemer, Kiernan, Essex, & Kupfer, 2008; Mitchell & Maxwell, 2013). Lagged mediator variables (m_{t-1} with t indicating the measurement occasion) are often used to "exogenize" the mediator. That is, using $m_{t-1} \rightarrow y_t$ instead of $m_t \rightarrow y_t$ to estimate the mediator-outcome part of the mediation model is assumed to solve issues of reverse causation biases and confounding. However, it is important to realize that simply measuring the mediator earlier in time than the outcome does neither guarantee a causal model of the form $m \rightarrow y$, nor does it lead to an unbiased estimate for the lagged effect $m_{t-1} \rightarrow y_t$. Regressing y_t on m_{t-1} in the presence of an unconsidered confounder (u) opens a "back-door channel" through $u_{t-1} \rightarrow m_{t-1}$ and $u_{t-1} \rightarrow u_t \rightarrow y_t$ (Bellemare, Masaki, & Pepinsky, 2017). Similarly, when one directly compares the two cross-lagged paths $m_{t-1} \rightarrow y_t$ and $y_{t-1} \rightarrow m_t$, one path can be more biased toward zero than the other, which opens the door to potential reverse causation biases. In other words, temporal precedence of variables turns the causal ("selection on observables") assumption of cross-sectional designs into a "no dynamics among unobservables" assumption. Both sets of assumptions are untestable with standard statistical mediation analysis. However, DDA can provide information in both cross-sectional and longitudinal data settings: First, DDA can be used to test causal effect directionality and unconfoundedness of the contemporaneous mediator-outcome relation while potentially adjusting for autoregressive and cross-lagged effects. Second, one can use DDA to test for unconfoundedness when estimating the cross-lagged paths $m_{t-1} \rightarrow y_t$ and $y_{t-1} \rightarrow m_t$. Such a modeling strategy constitutes a promising alternative to test the robustness of the proposed mediational mechanism teachers' behavioral skills $\rightarrow$ student prosocial behavior $\rightarrow$ student academic outcome.

In conclusion, DDA provides a promising new method for understanding mechanisms of change in developmental research. The findings from this study indicated that outcomes on student academic competence are mediated by student improvements in prosocial behavior. Thus, prosocial behaviors such as effective problem solving and social skills act as academic enablers (Elliott et al., 2004), meaning that increasing these skills in children will lead to improvements in their academic performance. IY TCM had a main effect on prosocial behaviors in the prior study (Reinke et al., 2018). Here, we now have evidence that prosocial behaviors lead to academic competence. Thus, if one were to dismantle the complex and multifaceted components of IY TCM, direct instruction of prosocial behaviors should be considered a critical feature of the intervention. Future studies using DDA with other complex school-based interventions can lead to refinement and improvement of these interventions and improved outcomes for children.

References

Back, L. T., Polk, E., Keys, C. B., & McMahon, S. D. (2016). Classroom management, school staff relations, school climate, and academic achievement: Testing a model with urban high schools. *Learning Environments Research, 19,* 397–410. http://dx.doi.org/10.1007/s10984-016-9213-x

Baker-Henningham, H., Scott, S., Jones, K., & Walker, S. (2012). Reducing child conduct problems and promoting social skills in a middle-income country: Cluster randomised controlled trial. *The British Journal of Psychiatry, 201,* 101–108. http://dx.doi.org/10.1192/bjp.bp.111.096834

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Beebee, H., Hitchcock, C., & Menzies, P. (2009). *The Oxford handbook of causation*. Oxford, England: Oxford University Press.

Bellemare, M. F., Masaki, T., & Pepinsky, T. B. (2017). Lagged explanatory variables and the estimation of causal effect. *The Journal of Politics, 79,* 949–963. http://dx.doi.org/10.1086/690946

Blueprints for Healthy Youth Development. (n.d.). *Incredible Years - Teacher Classroom Management*. Retrieved from https://www.blueprintsprograms.org/programs/699999999/incredible-years-teacher-classroom-management

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics, 130*(5), e1136–e1145. http://dx.doi.org/10.1542/peds.2012-0243

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica, 47,* 1287–1294. http://dx.doi.org/10.2307/1911963

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology, 98,* 550–558. http://dx.doi.org/10.1037/a0018933

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82,* 405–432. http://dx.doi.org/10.1111/j.1467-8624.2010.01564.x

Edossa, A. K., Schroeders, U., Weinert, S., & Artelt, C. (2018). The development of emotional and behavioral self-regulation and their effects on academic achievement in childhood. *International Journal of Behavioral Development, 42,* 192–202. http://dx.doi.org/10.1177/0165025416687412

Elliott, S. N., DiPerna, J. C., Mroch, A. A., & Lang, S. C. (2004). Prevalence and patterns of academic enabling behaviors: An analysis of teachers' and students' ratings for a national sample of learners. *School Psychology Review, 33,* 297–304.

Embry, D. D. (2002). The Good Behavior Game: A best practice candidate as a universal behavioral vaccine. *Clinical Child and Family Psychology Review, 5,* 273–297. http://dx.doi.org/10.1023/A:1020977107086

Ford, T., Hayes, R., Byford, S., Edwards, V., Fletcher, M., Logan, S., . . . Ukoumunne, O. C. (2019). The effectiveness and cost-effectiveness of the Incredible Years Teacher Classroom Management programme in primary school children: Results of the STARS cluster randomised controlled trial. *Psychological Medicine, 49,* 828–842. http://dx.doi.org/10.1017/S0033291718001484

Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology, 46,* 1454–1480. http://dx.doi.org/10.1037/a0020204

Fuller, W. A. (1987). *Measurement error models*. New York, NY: Wiley. http://dx.doi.org/10.1002/9780470316665

Gerbino, M., Zuffianò, A., Eisenberg, N., Castellani, V., Luengo Kanacri, B. P., Pastorelli, C., & Caprara, G. V. (2018). Adolescents' prosocial behavior predicts good grades beyond intelligence and personality traits. *Journal of Personality, 86,* 247–260. http://dx.doi.org/10.1111/jopy.12309

Gifford-Smith, M. (2000). *Teacher Social Competence Scale* (Fast Track Project Technical Report). Durham, NC: Duke University.

Gottfredson, D. C., Cook, T. D., Gardner, F. E., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science, 16,* 893–926. http://dx.doi.org/10.1007/s11121-015-0555-x

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., & Smola, A. J. (2008). A kernel statistical test of independence. *Advances in Neural Information Processing Systems, 20,* 585–592.

Hayes, R., Titheradge, D., Allen, K., Allwood, M., Byford, S., Edwards, V., . . . Russell, A. E. (2020). The Incredible Years Teacher Classroom Management programme and its impact on teachers' professional self-efficacy, work-related stress, and general well-being: Results from the STARS randomized controlled trial. *British Journal of Educational Psychology, 90,* 330–348. http://dx.doi.org/10.1111/bjep.12284

Herman, K. C., Reinke, W. M., & Dong, N. (2018, July)). *Group randomized evaluation of a classroom management program for middle school teachers*. Paper presented at the International School Psychology Association Conference, Tokyo, Japan.

Holmbeck, G. N., Franks Bruno, E., & Jandasek, B. (2006). Longitudinal research in pediatric psychology: An introduction to the special issue. *Journal of Pediatric Psychology, 31,* 995–1001. http://dx.doi.org/10.1093/jpepsy/jsj070

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *Journal of Experimental Education, 84,* 175–196. http://dx.doi.org/10.1080/00220973.2014.952397

Huang, F. L. (2018). Multilevel modeling and ordinary least squares regression: How comparable are they? *Journal of Experimental Education, 86,* 265–281. http://dx.doi.org/10.1080/00220973.2016.1277339

Hutchings, J., Martin-Forbes, P., Daley, D., & Williams, M. E. (2013). A randomized controlled trial of the impact of a teacher classroom management program on the classroom behavior of children with and without behavior problems. *Journal of School Psychology, 51,* 571–585. http://dx.doi.org/10.1016/j.jsp.2013.08.001

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15,* 309–334. http://dx.doi.org/10.1037/a0020761

Imai, K., Tingley, D., & Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society, 176,* 5–51. http://dx.doi.org/10.1111/j.1467-985X.2012.01032.x

Kellam, S. G., Mackenzie, A. C., Brown, C. H., Poduska, J. M., Wang, W., Petras, H., & Wilcox, H. C. (2011). The good behavior game and the future of prevention and treatment. *Addiction Science & Clinical Practice, 6,* 73–84.

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2009). Teacher Observation of Classroom Adaptation-Checklist: Development and factor structure. *Measurement and Evaluation in Counseling & Development, 42,* 15–30. http://dx.doi.org/10.1177/0748175609333560

Kraemer, H. C., Kiernan, M., Essex, M., & Kupfer, D. J. (2008). How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. *Health Psychology, 27,* S101–S108. http://dx.doi.org/10.1037/0278-6133.27.2(Suppl.).S101

Kremer, K. P., Flower, A., Huang, J., & Vaughn, M. G. (2016). Behavior problems and children's academic achievement: A test of growth-curve models with gender and racial differences. *Children and Youth Services Review, 67,* 95–104. http://dx.doi.org/10.1016/j.childyouth.2016.06.003

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36,* 249–277. http://dx.doi.org/10.1207/S15327906MBR3602_06

Larzelere, R. E., Gunnoe, M. L., Ferguson, C. J., & Roberts, M. W. (2019). The insufficiency of the evidence used to categorically oppose spanking and its implications for families and psychological science: Comment on Gershoff et al. *American Psychologist, 74,* 497–499. http://dx.doi.org/10.1037/amp0000461

Lewis, D. (1973). Causation. *The Journal of Philosophy, 70,* 556–567. http://dx.doi.org/10.2307/2025310

Li, X., & Wiedermann, W. (2019). Conditional direction dependence analysis: Evaluating the causal direction of effects in linear models with interaction terms. *Multivariate Behavioral Research*. Advance online publication. http://dx.doi.org/10.1080/00273171.2019.1687276

Malecki, C. K., & Elliot, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly, 17,* 1–23. http://dx.doi.org/10.1521/scpq.17.1.1.19902

McCormick, M. P., Cappella, E., O'Connor, E. E., & McClowry, S. G. (2015). Social-emotional learning and academic achievement: Using causal methods to explore classroom-level mechanisms. *AERA Open, 1,* 1–26. http://dx.doi.org/10.1177/2332858415603959

McNeish, D., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods, 24,* 20–35. http://dx.doi.org/10.1037/met0000182

Mitchell, M. A., & Maxwell, S. E. (2013). A comparison of the cross-sectional and sequential designs when assessing longitudinal mediation. *Multivariate Behavioral Research, 48,* 301–339. http://dx.doi.org/10.1080/00273171.2013.784696

Murray, D. W., Rabiner, D. L., Kuhn, L., Pan, Y., & Sabet, R. F. (2018). Investigating teacher and student effects of the Incredible Years Classroom Management Program in early elementary school. *Journal of School Psychology, 67,* 119–133. http://dx.doi.org/10.1016/j.jsp.2017.10.004

Nye, E., Melendez-Torres, G. J., & Gardner, F. (2018). Mixed methods systematic review on effectiveness and experiences of the Incredible Years Teacher Classroom Management programme. *Review of Education, 7,* 631–669. http://dx.doi.org/10.1002/rev3.3145

Office of Juvenile Justice and Delinquency Prevention's Model Programs. (2018). Program profile: Incredible Years Teacher Classroom Management program. Retrieved from https://www.crimesolutions.gov/ProgramDetails.aspx?ID=587

Patterson, G., Reid, J., & Dishion, T. (1992). *Antisocial boys*. Eugene, OR: Castalia Publishing Company.

Pianta, R. C. (1999). *Enhancing relationships between children and teachers*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10314-000

Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology, 66,* 29–38. http://dx.doi.org/10.1016/j.jesp.2015.09.012

Pituch, K. A., & Stapleton, L. M. (2012). Distinguishing between cross- and cluster-level mediation processes in the cluster randomized trial. *Sociological Methods & Research, 41,* 630–670. http://dx.doi.org/10.1177/0049124112460380

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879–903. http://dx.doi.org/10.1037/0021-9010.88.5.879

Rabiner, D. L., Godwin, J., & Dodge, K. A. (2016). Predicting academic achievement and attainment: The contribution of early academic skills, attention difficulties, and social competence. *School Psychology Review, 45,* 250–267. http://dx.doi.org/10.17105/SPR45-2.250-267

Raudenbush, S. (1989). Centering predictors in multilevel analysis: Choices and consequences. *Multilevel Modelling Newsletter, 1,* 10–12.

R Core Team. (2019). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reinke, W. M., Herman, K. C., & Dong, N. (2018). The incredible years teacher classroom management program: Outcomes from a group randomized trial. *Prevention Science, 19,* 1043–1054. http://dx.doi.org/10.1007/s11121-018-0932-3

Reinke, W. M., Herman, K. C., Petras, H., & Ialongo, N. S. (2008). Empirically derived subtypes of child academic and behavior problems: Co-occurrence and distal outcomes. *Journal of Abnormal Child Psychology, 36,* 759–770. http://dx.doi.org/10.1007/s10802-007-9208-2

Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly, 26,* 1–13. http://dx.doi.org/10.1037/a0022714

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104,* 700–712. http://dx.doi.org/10.1037/a0027268

Rosenström, T., & García-Velázquez, R. (2020). Distribution-based causal inference: A review and practical guidance for epidemiologists. In W. Wiedermann, D. Kim, E. Sungur, & A. von Eye (Eds.), *Direction dependence in statistical models: Methods of analysis* (pp. 267–294). Hoboken, NJ: Wiley.

Sawyer, A. C. P., Chittleborough, C. R., Mittinty, M. N., Miller-Lewis, L. R., Sawyer, M. G., Sullivan, T., & Lynch, J. W. (2015). Are trajectories of self-regulation abilities from ages 2–3 to 6–7 associated with academic achievement in the early school years? *Child: Care, Health and Development, 41,* 744–754. http://dx.doi.org/10.1111/cch.12208

Sayal, K., Washbrook, E., & Propper, C. (2015). Childhood behavior problems and academic outcomes in adolescence: Longitudinal population-based study. *Journal of the American Academy of Child & Adolescent Psychiatry, 54,* 360–368. http://dx.doi.org/10.1016/j.jaac.2015.02.007

Schaeffer, C. M., Petras, H., Ialongo, N., Poduska, J., & Kellam, S. (2003). Modeling growth in boys' aggressive behavior across elementary school: Links to later criminal involvement, conduct disorder, and antisocial personality disorder. *Developmental Psychology, 39,* 1020–1035. http://dx.doi.org/10.1037/0012-1649.39.6.1020

U.S. Department of Education. (2014). *What Works Clearinghouse: Procedures and standards handbook* (Version 3.0). Retrieved from https://ies.ed.gov/ncee/wwc/

von Eye, A., & DeShon, R. P. (2012). Directional dependence in developmental research. *International Journal of Behavioral Development, 36,* 303–312. http://dx.doi.org/10.1177/0165025412439968

Webster-Stratton, C. (2018). The Incredible Years Teacher Classroom Management program. Retrieved from www.incredibleyears.com/download/administrators/TCM-logic-model.pdf

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in head start. *Journal of Clinical Child Psychology, 30,* 283–302. http://dx.doi.org/10.1207/S15374424JCCP3003_2

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology, 33,* 105–124. http://dx.doi.org/10.1207/S15374424JCCP3301_11

Webster-Stratton, C., Reid, M. J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years Teacher and Child Training Programs in high-risk schools. *Journal of Child Psychology and Psychiatry, 49,* 471–488. http://dx.doi.org/10.1111/j.1469-7610.2007.01861.x

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology, 85,* 357–364. http://dx.doi.org/10.1037/0022-0663.85.2.357

Wiedermann, W., & Li, X. (2018). Direction dependence analysis: A framework to test the direction of effects in linear models with an implementation in SPSS. *Behavior Research Methods, 50,* 1581–1601. http://dx.doi.org/10.3758/s13428-018-1031-x

Wiedermann, W., & Li, X. (2019). Direction dependence analysis. Retrieved from www.ddaproject.com

Wiedermann, W., Li, X., & von Eye, A. (2019). Testing the causal direction of mediation effects in randomized intervention studies. *Prevention Science, 20,* 419–430. http://dx.doi.org/10.1007/s11121-018-0900-y

Wiedermann, W., Merkle, E. C., & von Eye, A. (2018). Direction of dependence in measurement error models. *British Journal of Mathematical and Statistical Psychology, 71,* 117–145. http://dx.doi.org/10.1111/bmsp.12111

Wiedermann, W., & Sebastian, J. (2019a). Direction dependence analysis in the presence of confounders: Applications to linear mediation models using observational data. *Multivariate Behavioral Research*. Advance online publication. http://dx.doi.org/10.1080/00273171.2018.1528542

Wiedermann, W., & Sebastian, J. (2019b). Sensitivity analysis and extensions of testing the causal direction of dependence: A rejoinder to Thoemmes (2019). *Multivariate Behavioral Research*. Advance online publication. http://dx.doi.org/10.1080/00273171.2019.1659127

Wiedermann, W., & von Eye, A. (2015). Direction of effects in mediation analysis. *Psychological Methods, 20,* 221–244. http://dx.doi.org/10.1037/met0000027

Wiedermann, W., & von Eye, A. (2016). Directionality of effects in causal mediation analysis. In W. Wiedermann & A. von Eye (Eds.), *Statistics and causality: Methods for applied empirical research* (pp. 63–106). Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9781118947074.ch4

Zima, B. T., Hurlburt, M. S., Knapp, P., Ladd, H., Tang, L., Duan, N., . . . Wells, K. B. (2005). Quality of publicly-funded outpatient specialty mental health care for common childhood psychiatric disorders in California. *Journal of the American Academy of Child & Adolescent Psychiatry, 44,* 130–144. http://dx.doi.org/10.1097/00004583-200502000-00005

Received September 22, 2019
Revision received March 30, 2020
Accepted June 9, 2020 ■