



# Implementation Matters: Using Complier Average Causal Effect Estimation to Determine the Impact of the Promoting Alternative Thinking Strategies (PATHS) Curriculum on Children's Quality of Life

DOI:

[10.1037/edu0000360](https://doi.org/10.1037/edu0000360)

## Document Version

Final published version

[Link to publication record in Manchester Research Explorer](#)

## Citation for published version (APA):

Panayiotou, M., Humphrey, N., & Hennessey, A. (2019). Implementation Matters: Using Complier Average Causal Effect Estimation to Determine the Impact of the Promoting Alternative Thinking Strategies (PATHS) Curriculum on Children's Quality of Life. *Journal of Educational Psychology*, 112(2), 236-253. Advance online publication. <https://doi.org/10.1037/edu0000360>

## Published in:

Journal of Educational Psychology

## Citing this paper

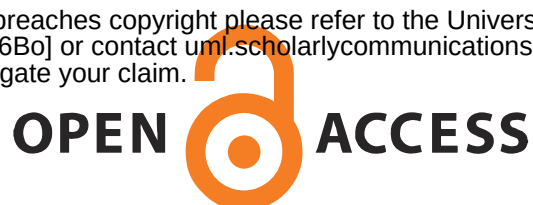
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

## General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

## Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [um.scholarlycommunications@manchester.ac.uk](mailto:um.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



# Journal of Educational Psychology

## **Implementation Matters: Using Complier Average Causal Effect Estimation to Determine the Impact of the Promoting Alternative Thinking Strategies (PATHS) Curriculum on Children's Quality of Life**

Margarita Panayiotou, Neil Humphrey, and Alexandra Hennessey

Online First Publication, April 8, 2019. <http://dx.doi.org/10.1037/edu0000360>

### CITATION

Panayiotou, M., Humphrey, N., & Hennessey, A. (2019, April 8). Implementation Matters: Using Complier Average Causal Effect Estimation to Determine the Impact of the Promoting Alternative Thinking Strategies (PATHS) Curriculum on Children's Quality of Life. *Journal of Educational Psychology*. Advance online publication. <http://dx.doi.org/10.1037/edu0000360>

# Implementation Matters: Using Complier Average Causal Effect Estimation to Determine the Impact of the Promoting Alternative Thinking Strategies (PATHS) Curriculum on Children's Quality of Life

Margarita Panayiotou, Neil Humphrey, and Alexandra Hennessey  
The University of Manchester

This cluster randomized trial evaluated the impact of the Promoting Alternative Thinking Strategies (PATHS) curriculum on children's psychological wellbeing, peer social support, and school connectedness. Forty-five schools in England were randomly assigned to implement PATHS or continue their usual provision for 2 years. The trial sample was 5,218 students, aged 7–9 at baseline ( $M = 8.12$ ,  $SD = .87$ ). Teachers in PATHS schools received initial training and ongoing implementation support and assistance from trained coaches. Multilevel intent-to-treat (ITT) analysis of outcome data indicated that PATHS led to a small, statistically significant improvement in children's psychological wellbeing but had no discernible impact on their peer social support or school connectedness. Multilevel complier average causal effect estimation using dosage as a compliance marker increased the intervention effect size for psychological wellbeing and revealed significant medium to large effects for peer social support and school connectedness. The implications of these findings are discussed, and study limitations are noted.

### *Educational Impact and Implications Statement*

This study demonstrates that the Promoting Alternative Thinking Strategies (PATHS) curriculum significantly improves children's psychological wellbeing, but also social peer support and school connectedness, once intervention compliance is considered. Thus, social and emotional learning interventions such as PATHS can be considered an efficacious means for promoting these outcomes, but further efforts to optimize implementation are required if their true potential is to be realized.

*Keywords:* social and emotional learning, intervention, implementation, randomized trial, complier average causal effect

*Supplemental materials:* <http://dx.doi.org/10.1037/edu0000360.supp>

Universal, school-based social and emotional learning (SEL) interventions aim to promote children's skills of social and emotional competence including their self-awareness, self-management, social awareness, relationship skills, and responsible decision-making, through explicit instruction in the context of safe, caring, well-managed, and participatory learning environments (CASEL, 2003; Weissberg, Durlak, Domitrovich, & Gullotta, 2015). Five major meta-analyses have rigorously demonstrated that SEL interventions can lead to significant and meaningful short-term (Corcoran,

Cheung, Kim, & Xie, 2018; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Sklad, Diekstra, Ritter, Ben, & Gravestijn, 2012; Wigelsworth et al., 2016) and long-term (Taylor, Oberle, Durlak, & Weissberg, 2017) improvements in a range of student outcomes, including social and emotional competence, mental health difficulties, and academic attainment. Average effect sizes for these outcomes compare very favorably to other forms of universal preventive interventions. For example, the mean SEL effect size of 0.57 for social and emotional competence reported by Durlak et al. (2011) is at the 75th percentile in the empirical distribution of prevention program effects (Tanner-Smith, Durlak, & Marx, 2018).

However, the majority (66%) of trials published to date have been led by (or involved) program developers (Wigelsworth et al., 2016), which may amplify intervention effects as a result of higher quality implementation, bias, or both; independent replication is therefore essential (Open Science Collaboration, 2015). In addition, SEL intervention research has neglected some outcome domains, despite their importance in underpinning theory. Specifically, the impact of SEL programs on children's psychological wellbeing, school connectedness, and perceptions of peer relations

---

Margarita Panayiotou, Neil Humphrey, and Alexandra Hennessey, Manchester Institute of Education, The University of Manchester.

This study was part of a research project funded by the National Institute for Health Research.

We thank Linda K. Muthén and Booil Jo for providing insightful comments on preliminary analyses.

Correspondence concerning this article should be addressed to Neil Humphrey, Manchester Institute of Education, The University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom. E-mail: [neil.humphrey@manchester.ac.uk](mailto:neil.humphrey@manchester.ac.uk)

and social support remains underexplored (Humphrey, Hennessey et al., 2018). Finally, although we know that variability in the implementation of school-based interventions is inevitable (Durlak, 2016), most trials fail to properly account for this fact in their analyses. Thus, there are increasing calls to move beyond intent-to-treat (ITT) analyses in randomized trials in education, which assume complete compliance among those allocated to the intervention group (Peugh, Strotman, McGrady, Rausch, & Kashikar-Zuck, 2017). The current study, therefore, showcases independent research on an evidence-based SEL intervention (the Promoting Alternative Thinking Strategies [PATHS] curriculum; Kusche & Greenberg, 1994), focusing on its impact on theoretically plausible but hitherto neglected outcome domains, and taking account of implementation variability by contrasting ITT findings with those produced through complier average causal effect (CACE) estimation. We begin with a tripartite exposition of our trial outcomes.

### **Psychological Wellbeing, Peer Social Support, and School Connectedness: Conceptualization and Developmental Significance**

Wellbeing is an important constituent component in the complete state model of mental health (Keyes, 2005) and is a key area of focus in work on children's rights, early learning, school-based program development, and public policy (Bonell et al., 2014; Mashford-Scott, Church, & Tayler, 2012; World Health Organization, 2003; Young Minds & National Children's Bureau, 2017). Although exact definitions vary according to the degree of emphasis placed on hedonic and eudaimonic aspects, and the relative importance of objective versus subjective approaches to measurement, there is broad agreement that it refers to our sense of "how well we are, and how our lives are going" (Children's Society, 2017, p. 7). Psychological wellbeing can thus be conceptualized as comprising elements such as positive affect, life satisfaction, and self-acceptance (Keyes, 2006; Ravens-Sieberer et al., 2007). Its developmental significance is evident in studies that demonstrate prospective relations with academic attainment (Gutman & Vorhaus, 2012), employment (Layard, Clark, Cornaglia, Powdthavee, & Verhoef, 2014), and even mortality (Chida & Steptoe, 2008). That is, children who experience higher levels of psychological wellbeing are more likely to do well in school, be employed as adults, and live longer.

Peer social support is "a system of giving and receiving help founded on key principles of respect, shared responsibility, and mutual agreement of what is helpful . . . and about understanding another's situation empathically" (Mead, Hilton, & Curtis, 2001, p. 135). This system increases in importance during the course of development. In middle childhood, more than 30% of children's social interactions involve their peers (Gifford-Smith & Brownell, 2003) and by adolescence a substantial amount of their time is spent within the peer context (Monahan, Guyer, Silk, Fitzwater, & Steinberg, 2016). Peers serve as both social (Crosnoe, Cavanagh, & Elder, 2003) and academic resources (e.g., by assisting one another, sharing notes and books; Wentzel, 1991). They can provide a sense of security, acceptance, emotional support, and opportunities to learn skills (Kochel, Bagwell, Ladd, & Rudolph, 2017). Supportive peer relations can help children be resilient to feelings of loneliness and depression, reduce the risk for peer victimization, and contribute to positive school adjustment (Bag-

well & Schmidt, 2011; Kendrick, Jutengren, & Stattin, 2012; Kochel et al., 2017; Parker & Asher, 1993).

School connectedness has been defined in various ways (Panayiotou, Humphrey, & Wigelsworth, 2019) with SEL researchers tending to focus on specific constructs such as school engagement or attachment (e.g., CASEL, 2003; Ross & Tolan, 2018) or use the umbrella term of "school attitudes" (Durlak et al., 2011; Taylor et al., 2017; Zins, Bloodworth, Weissberg, & Walberg, 2004). It can be broadly conceived of as relating to affective (e.g., feelings about the school, teachers, and/or peers), cognitive (e.g., perceptions and beliefs related to self, school, and peers), and behavioral (e.g., academic engagement, participation in extracurricular activities) aspects of students' school experience (Jimerson, Campos, & Greif, 2003; Libbey, 2004). In terms of developmental significance, school connectedness has been shown to be associated with academic achievement (LeCroy & Krysik, 2008; Li & Lerner, 2011; Wang & Holcombe, 2010), general healthy development, and less mental health difficulties and risky behaviors (Bond et al., 2007; Catalano, Haggerty, Oesterle, Fleming, & Hawkins, 2004; Henry & Slater, 2007; Kia-Keating & Ellis, 2007; Li & Lerner, 2011; Panayiotou et al., 2019).

Collectively then, our three focal outcomes can be conceptualized as important elements of children's quality of life (Ravens-Sieberer et al., 2007). In the following section we consider the mechanisms by which they may be improved through systematic exposure to a universal SEL intervention.

### **Social and Emotional Learning: Proposed PATHS to Improved Quality of Life**

The aim of the PATHS curriculum is to promote self-control, emotional understanding, positive self-esteem, relationships, and interpersonal problem-solving skills among children in preschool and elementary education settings. It is one of the most well-validated SEL interventions, with numerous randomized controlled trials (RCTs) spanning the United States (Conduct Problems Prevention Research Group, 1999a, 1999b; Domitrovich, Cortes, & Greenberg, 2007; Greenberg, Kusche, Cook, & Quamma, 1995; Morris et al., 2014; Schonfeld et al., 2015; Social Character Development Research Consortium, 2010), United Kingdom (Berry et al., 2016; Ross, Sheard, Cheung, Elliott, & Slavin, 2011), Switzerland (Malti, Ribeaud, & Eisner, 2011), and Croatia (Novak, Mihić, Bašić, & Nix, 2017). In elaborating the hypothesized mechanisms by which the PATHS curriculum could impact upon children's psychological wellbeing, peer social support, and school connectedness, we draw upon SEL theory (CASEL, 2003), the program logic model and theoretical rationale (Center for the Study and Prevention of Violence, 2012; Greenberg & Kusche, 1993) and the extant evidence base (e.g., Oberle, 2018).

Turning first to psychological wellbeing, we theorized that improvements would be driven directly by improved proficiency in the SEL domains that are central to, and taught explicitly in, programs such as the PATHS curriculum (Bywater & Sharples, 2012; Humphrey, Hennessey et al., 2018). Indeed, these have been described as, "the skills and competencies that underlie mental health" (Weare & Markham, 2005, p. 14). Children who are able to understand, articulate, and manage their emotions, while also being better equipped to develop and maintain positive social relationships (including social problem solving) are more likely to

experience greater levels of positive affect: “Emotions can need regulating when they threaten to overwhelm or need to be amplified . . . these [social-emotional] skills help them to experience more well-being and maintain satisfying relationships with others” (Denham, 2006, p. 70). Research on the determinants of wellbeing provides support for these propositions. For example, several studies have evidenced positive associations between aspects of social-emotional competence and wellbeing (Brackett & Mayer, 2003; Jones, Greenberg, & Crowley, 2015; Mavroveli, Petrides, Rieffe, & Bakker, 2007; Moffitt et al., 2011). More generally, research has also found that a supportive classroom climate predicts levels of youth wellbeing (Oberle, 2018). Finally, Ashdown and Bernard’s (2012) RCT of the You Can Do It SEL intervention led to significant improvements in the wellbeing of students in early childhood educational contexts.

With regard to peer social support, we note that improved prosocial peer relations are indicated as a proximal outcome in the PATHS program logic model (Center for the Study and Prevention of Violence, 2012). We hypothesized that the relational and prosocial foci of the intervention (e.g., lessons on getting along with others, being responsible and caring for others, making and keeping friends) would plausibly lead to children being better equipped to develop and maintain cooperative social relationships with their peers, and in turn recognize the support they receive from them. We further theorized that this would be reinforced by concurrent improvements in social awareness through explicit instruction focusing on empathy and perspective-taking. As above, there is some empirical support for these propositions. For example, we know that early social-emotional competence acts as a foundation of later social competence and successful peer interaction and peer-rated likability (Denham et al., 2003; Hay, Payne, & Chadwick, 2004). Similarly, a recent meta-analysis found that the ability to consider other people’s thoughts and feelings is directly related to children’s tendency to act prosocially (Imuta, Henry, Slaughter, Selcuk, & Ruffman, 2016).

Finally, SEL theory forecasts a stronger sense of school connectedness among children exposed to programs such as PATHS because they create learning environments that are well-managed and participatory, and in which students feel cared for and safe (CASEL, 2003). Here we can draw parallels with attachment theory, in which the development of secure attachment representations (e.g., “I am good, deserving and capable;” “Others are available and responsive to my needs;” “The world is safe”) is supported through the caregiver’s provision of a “secure base and safe haven” characterized by warmth, emotional support, and stimulation (Zeanah, Berlin, & Boris, 2011). As children become more autonomous, meaningful connections with the schools they attend can be formed, given similar conditions (Oldfield, Humphrey, & Hebron, 2016). Indeed, evidence suggests that students in fair and supportive environments are more likely to bond with their school (Roeser, van der Wolf, & Strobel, 2001), unlike those who are victimized and treated unfairly (Ripski & Gregory, 2009). Research has also demonstrated that students’ social-emotional competence is prospectively associated with their school connectedness (Panayiotou et al., 2019; Ross & Tolan, 2018).

Thus, extant theory and evidence support the proposition that the PATHS curriculum could feasibly lead to meaningful improvements in children’s psychological wellbeing, peer social support, and school connectedness. However, research in implementation

science consistently finds that one of the most important factors influencing the achievement of intervention outcomes is the level of implementation that is achieved (Durlak, 2016). It is this issue to which we now turn.

## Implementation Matters

*Implementation* refers to “the process of putting a practice or program into place” (Forman, 2015, p. 10). Despite a significant rise in the proportion of SEL trials reporting on implementation in the last two decades (from just 5% in an early review to 69% in a recent meta-analysis; Durlak, 1997; Wigelsworth et al., 2016), most offer only descriptive data (e.g., proportion of sessions delivered, average fidelity ratings). This provides evidence of the degree to which a given intervention was actually delivered but represents a significant missed opportunity in terms of use of such data to support estimation of effects that take implementation variability into account (Berg, Bradshaw, Jo, & Jalongo, 2017).

Even with successful randomization, the estimation of intervention effects can be biased unless everyone complies with the intervention (Jo & Muthén, 2001). Theory and research in implementation science suggest that variability in delivery of school-based interventions is inevitable, and SEL programs are no exception (Durlak, 2016). This variability is assumed to influence program outcomes (Durlak & DuPre, 2008; Durlak et al., 2011). PATHS specifically is a “spiral curriculum,” in which topics and concepts are revisited; units and lessons are developmentally sequenced; new learning is linked to previous learning; and the competence of learners increases with each successive visit to a topic or concept (Humphrey et al., 2015). Given this, higher levels of exposure are arguably highly likely to underpin amplified intervention outcomes because of the increased opportunities afforded for reinforcement, consolidation, and generalization of learning.

Traditional approaches to analyzing the relationship between implementation and outcomes (e.g., “as treated” and “per protocol”) have been found wanting, as they introduce even more bias by stripping out data from noncompliers (Sedgwick, 2015). CACE and related instrumental variable approaches overcome this problem by using data from compliers and noncompliers across the intervention and control arms of a trial, thereby producing an unbiased intervention effect estimate (Peugh et al., 2017). However, despite being commonplace in other fields (e.g., medicine; Stuart, Perry, Le, & Jalongo, 2008), they have been given “little to no attention in school psychology” (Peugh & Toland, 2017, p. 5). Indeed, following a systematic literature search undertaken in preparation for the current paper, we found just 11 published examples of the application of CACE in the context of a RCT of any school-based intervention (Berg et al., 2017; Cogo-Moreira, Brandao de Ávila, Ploubidis, & Mari, 2013; Connell, 2009; Connell, Dishion, Yasui, & Kavanagh, 2007; Fosco, Van Ryzin, Connell, & Stormshak, 2016; Jago et al., 2015; Schultz, Evans, Langberg, & Schoemann, 2017; Stormshak et al., 2011; Stuart et al., 2008; Van Ryzin, Stormshak, & Dishion, 2012; Véronneau, Dishion, Connell, & Kavanagh, 2016). Despite the successful application of CACE in each of the above studies, compliance was examined at the individual level, which could lead to variance misestimation in cases where the effects of clustering and non-compliance are not—but should be—considered together (Jo,

Asparouhov, Muthén, Ialongo, & Brown, 2008). To date, the application of CACE in a multilevel context has been scarce. Therefore, we sought to address this crucial gap in the literature.

### The Current Study

The current study reports on a major RCT of the PATHS curriculum, commissioned by the National Institute for Health Research in England (Ref: 10/3006/01; Humphrey, Hennessey et al., 2018). To date, other study outputs have been produced relating to the impact of PATHS on children's social and emotional competence, mental health difficulties (Humphrey, Barlow et al., 2016) and academic progress (Humphrey et al., 2015). Some of these outcomes appeared to vary as a function of levels of implementation (Humphrey, Barlow, & Lendrum, 2018; Humphrey, Hennessey et al., 2018). In this article, we turn to the impact of PATHS on children's psychological wellbeing, peer social support and school connectedness, with the secondary aim of determining the extent to which the magnitude of any intervention effects change once intervention compliance is taken into account. No previous study has focused on the impact of PATHS on the above noted outcomes, and although a minority have examined the moderating role of implementation variability (Berry et al., 2016; Conduct Problems Prevention Research Group, 1999b; Schonfeld et al., 2015), none have used CACE. Building on existing work (Jo, Asparouhov, Muthén, Ialongo, et al., 2008), one of our primary goals was, therefore, to provide an example of multilevel CACE analysis for a school-based cluster RCT, which has been notably absent from the existing literature (for Mplus code see online supplementary material file).

Based on the preceding theory and evidence, we predicted that intervention exposure would lead to significant improvements in students' psychological wellbeing (H1a), peer social support (H1b), and school connectedness (H1c) when compared to usual provision. However, we also predicted that the intervention effects noted above would be amplified once intervention compliance (using dosage as an appropriate marker) was taken into account (H2a–c).

### Method

#### Design

A two-group parallel cluster RCT was undertaken, with schools as the unit of randomization. Figure 1 depicts the flow of schools and children through the study. Random allocation of schools was performed independently of the authors by a local clinical trials unit, and was balanced by proportions of children eligible for free school meals (FSM) and speaking English as an additional language (EAL) via adaptive stratification (minimization).

Eligible schools were mainstream, state-maintained elementary schools in 10 local authorities (akin to school districts) in the north-west of England. Children attending participating schools in Years 3–5 (aged 7–9 years) were the target population. Participation required consent from the schools' head teachers. Child assent and parental opt-out consent were also sought. In total, 133 parents (2.5%) exercised their right to opt their children out of the trial, and no children declined assent or exercised their right to withdraw from the study. Baseline assessment of outcomes took place in

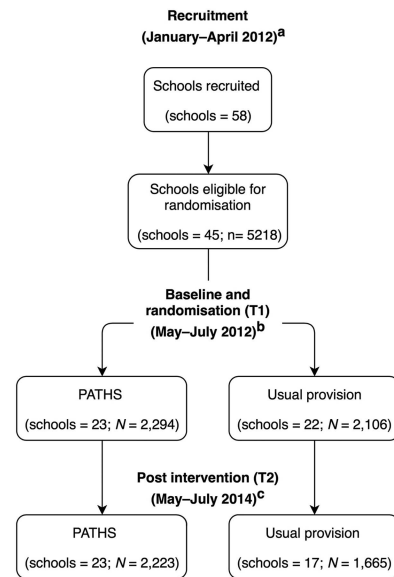


Figure 1. Flow of schools and students through the trial. <sup>a</sup> Thirteen schools failed to complete baseline measures because of lack of time ( $n = 7$ ), lack of information technology facilities ( $n = 2$ ), other priorities (e.g., Ofsted;  $n = 1$ ) and lack of response ( $n = 3$ ); <sup>b</sup> Student surveys were not completed because of absence, changing school, noncompletion or parental opt-out ( $n = 818$ ); <sup>c</sup> Of the students who provided baseline data,  $n = 420$  were lost to follow-up because of school-level attrition, and a further  $n = 623$  were lost to follow-up due to student attrition (e.g., absence, changing school or parental opt-out). However,  $n = 531$  who were part of the sample at baseline but provided data at T2 only were gained.

summer 2012, with follow-up in summer 2014. Assessment of implementation took place in the intervening period. The study received ethical approval from the ethics committee at the authors' host institution (Ref: 11,470).

#### Participants

**Schools.** Fifty-eight schools were initially recruited, out of which 45 met the eligibility criteria for randomization (completion of baseline measures and signing a memorandum of agreement to adhere to the trial protocol). Excluded schools were those who failed to complete baseline measures, most frequently citing a lack of time to do so (see Figure 1). Participating schools had significantly higher proportions of students eligible for FSM (29.7%) and speaking EAL (23.3%) than national averages (Department for Education, 2012) but were otherwise representative of norms in England in terms of size, attendance, attainment, ethnicity, and the proportion of students identified as having special educational needs (SEN; see Humphrey, Hennessey et al., 2018 for a description by assignment group).

**Students.** There were 5,218 students (male;  $n = 2,681$ , 51.4%) aged 7–9 years ( $M = 8.12$ ,  $SD = .87$ ) in trial schools at baseline. Their characteristics mirrored those of students in English elementary schools, albeit with a similar pattern of deviation in terms of FSM (29.4%), EAL (21.8%), and SEN (19%) to that noted above (Department for Education, 2012, 2013). Similar to the national picture of state-funded elementary schools (Depart-

ment for Education, 2012), 67.8% of the sample were Caucasian ( $n = 3,538$ ), 11.2% Asian ( $n = 585$ ), 7.1% Afro-Caribbean ( $n = 368$ ), 5.5% mixed race ( $n = 289$ ), 2.7% other/unclassified race ( $n = 143$ ), and 0.6% Chinese ( $n = 33$ ). Ethnicity data were not available for the remaining 262 (5%) students.

## Intervention

The theoretical underpinning of PATHS is the Affective-Behavioral-Cognitive-Developmental model of development (Greenberg & Kusche, 1993). The core component of the program is a taught curriculum, which is supplemented by generalization activities and techniques, and parent materials. Each class receives curriculum packs containing lessons and send-home activities built around four conceptual units (emotional understanding; self-control; social problem-solving; peer relations and self-esteem), plus associated material resources (e.g., posters, feelings dictionaries). Of particular additional note in the context of our psychological wellbeing hypothesis (H1) is the “PATHS Kid of the Day” procedure. Each day, a different focal child’s positive qualities and attributes are highlighted by other members of their class, their teacher, and parents/carers; doing so, we argue, will amplify feelings of self-acceptance and life satisfaction. In the current trial, class teachers also received a guidance manual developed by the research team that emphasized the PATHS program theory and the importance of effective implementation (available on request). Of particular note, this manual also contained a delivery schedule, which teachers were asked to follow. Given the spiral nature of the curriculum, delivering the lessons in the prespecified order was emphasized.

PATHS is implemented by class teachers as part of the general classroom timetable. In the current trial, all were qualified teachers (81% female) and had an average of 8 years teaching experience. PATHS curriculum packs contain an average of 40 lessons (30–40 min each) and are designed to be delivered throughout the school year, approximately twice per week. PATHS teachers received a full day of initial training with a half-day follow-up 4 months later. This training was led by certified trainers from Pennsylvania State University (PSU) and included a range of activities designed to familiarize teachers with PATHS theory, concepts and materials. PATHS teachers also received ongoing technical support and assistance (e.g., lesson modeling, observation and feedback) from three coaches, who were trained by PSU staff and received ongoing supervision throughout the trial.

Control schools reported implementing universal initiatives such as the social and emotional aspects of learning program, National Healthy Schools program, and Circle Time throughout the trial, in addition to personal, social, and health education as part of the standard school curriculum (see Humphrey, Hennessey et al., 2018 for more details).

## Measures

**Outcome variables.** The psychological wellbeing (seven items), peers and social support (four items), and school environment (four items) subscales of the child self-report version of the Kidscreen-27 (KS27; Ravens-Sieberer et al., 2007) were used. The psychological wellbeing domain of KS27 includes items on positive affect, satisfaction with life, and feeling emotionally balanced

(e.g., “Has your life been enjoyable?”). The peers and social support domain assesses the nature of the respondent’s relationships with their peers (e.g., “Have you been able to rely on your friends?”). Finally, the school environment domain explores feelings about school, learning and concentration, and perceptions of cognitive capacity (e.g., “Have you got on well with your teachers?”). Consistent with the broader literature on the measurement of school connectedness, this subscale measures dimensions of cognitive, behavioral, and affective school connectedness (Libbey, 2004).

In all three subscales, the respondent reads a statement and indicates their level of agreement on a 5-point scale ranging from 1 (*never*) to 5 (*always*). The KS27 is psychometrically robust, with high internal consistency ( $\alpha > .80$  for each subscale), clear factor structure with acceptable confirmatory factor analysis fit indices (comparative fit index = 0.96, root square mean of approximation = 0.07), good test–retest reliability (intra class correlation coefficients  $> .60$  for each subscale), and strong criterion validity (Ravens-Sieberer et al., 2007; Robitail, Robitail, Ravens-Sieberer et al., 2007).

**Covariates.** Conduct problems were measured by the relevant subscale of the teacher-report Strength and Difficulties Questionnaire (SDQ; Goodman, 1997), and social-emotional competence was assessed using the self-report Social Skills Improvement System (Gresham & Elliott, 2008). Both measures have been shown to be psychometrically robust (Gresham, Elliott, Vance, & Cook, 2011; Stone, Otten, Engels, Vermulst, & Janssens, 2010). Data on children’s sex and FSM eligibility were drawn from the National Pupil Database.

## Analytic Strategy

**CACE overview and assumptions.** In the intervention condition of any RCT, compliance can be observed and individuals can be categorized as compliers or noncompliers (Jo & Muthén, 2001). Assuming a binary treatment assignment ( $T_i$ ; 0 = control, 1 = treatment) and treatment received ( $D_i$ ; 0 = not received, 1 = received) for the individual  $i$ , Angrist, Imbens, and Rubin (1996) defined four possible participant behaviors, namely compliers, never-takers, defiers, and always-takers. *Compliers* are individuals who do what they are assigned to do ( $D_i = 1 \mid T_i = 1$  and  $D_i = 0 \mid T_i = 0$ ). *Never-takers* are those who do not receive the treatment regardless of assignment ( $D_i = 0 \mid T_i = 1$  and  $D_i = 0 \mid T_i = 0$ ). *Defiers* are participants who do the opposite of what they are expected to do ( $D_i = 0 \mid T_i = 1$  and  $D_i = 1 \mid T_i = 0$ ), and *always-takers* are those who always receive the treatment irrespective of their assigned condition ( $D_i = 1 \mid T_i = 1$  and  $D_i = 1 \mid T_i = 0$ ). Causal interpretation with CACE is possible assuming that the potential outcome of each individual is not affected by the treatment assignment of other individuals (i.e., stable unit treatment value [SUTVA]), treatment assignment is random, there are no defiers (i.e., monotonicity) or always-takers, and the outcome for never-takers is independent of the treatment assignment. In other words, the treatment effect is zero for those who did not participate (i.e., exclusion restriction assumption; see Angrist et al., 1996).

However, given that compliance status is usually unknown for those in the control group, CACE models are estimated probabilistically as structural equation mixture models (Peugh et al., 2017). In other words, CACE identifies those in the control group

that would have potentially complied with the intervention, had they been randomized to receive it. This is visually represented in Figure 2, where the circles represent latent variables, and the squares represent observed variables. The compliance status is observed (known) for the intervention group and latent (unknown) for the control group. The variable  $U$  in Figure 2 represents the observed binary compliance variable (based on dosage) for the intervention group, where 0 = noncompliers and 1 = compliers. Compliance data is treated as missing in  $U$  variable for the control group and it is therefore estimated through the categorical latent mixture variable  $C$ , which represents the compliance status (Class 1 = noncompliers; Class 2 = compliers). This is achieved using the observed compliance data in  $U$  that is available for the intervention group, the missing data in  $U$  for the control group, and the response variable distribution information for the sample (Peugh et al., 2017). Using mixture modeling, potential compliers from the control group (had they received the intervention) are therefore identified based on the similarities of their response scores to those of compliers from the intervention group (Peugh et al., 2017).

In Figure 2, the continuous dependent variable *outcome* is regressed on the student-level and school-level covariates and the binary treatment variable (PATHS vs. control). The arrows from covariates to  $C$  represent the multinomial logistic regression (in student-level) and linear regression (in school-level) of  $C$  on covariates (i.e., predictors of compliance). The arrow from  $C$  to outcome variable indicates that the intercept of outcome varies across the classes of  $C$  (compliers vs. noncompliers). The dashed lines in the student-level indicate that the residual variance of outcome and the regression slopes of outcome on covariates vary across the classes of  $C$ . The dashed lines in the school-level indicate that the residual variance of outcome, the slopes of outcome on covariates, and the slope of outcome on treatment variable (PATHS vs. control) vary across the classes of  $C$  (for further information on multilevel CACE estimation, see Jo, Asparouhov, Muthén et al., 2008).

**Defining compliance.** The current study uses dosage as a marker for intervention compliance because the primary motivation for the CACE parameter is to determine treatment effects in the context of receipt of an intervention (as opposed to the offer of

said intervention, as in ITT estimation). Dosage is arguably the best indicator of this, as the unit of measurement is the amount of the intervention received (Schochet & Chiang, 2011).

Dosage data were collected by three trained research assistants in structured observations (one per class per year) conducted in the autumn and winter terms (November to April) in a given school year as part of a comprehensive implementation and process evaluation (IPE; Humphrey, Hennessey et al., 2018). Use of independent observations to capture implementation data is considered much more robust than teacher self-report methods, which are subject to substantial positive bias (Hansen, 2014). A projected dosage indicator (% lessons delivered by the end of the school year) was recorded for the first (Time 1 [T1]) and second (Time 2 [T2]) years of the trial based on progress against the delivery schedule included in the aforementioned implementation manual. For example, if at the point of observation, the teacher should have been delivering lesson number 10 according to the delivery schedule, and they were in fact delivering Lesson 5, this would be coded as 50%. Similarly, if at the point of observation, the teacher should have been delivering Lesson 18 according to the delivery schedule, and they were in fact delivering Lesson 12, this would be coded as 67%.

Given that CACE requires a binary indicator of compliance, continuous implementation variables (variable  $U$  in Figure 2) are dichotomized into compliers (score of 1) and noncompliers (score of 0). However, in the absence of a verified compliance cut-off, as in the current study, there is a trade-off when defining compliance (Stuart et al., 2008). Lower cut-off values may lead to great variation in the implementation of the intervention within compliers (Berg et al., 2017). Higher cut-offs, on the other hand, may lead to larger CACE estimates but can also result in the exclusion restriction assumption being less realistic (Connell, 2009). In such cases, sensitivity analyses that compare the results of different compliance thresholds can be performed (Sagarin et al., 2014). Compliance in the current study was, therefore, defined in two ways (as in Berg et al., 2017). Classrooms that fell above the 50th percentile at the T1 implementation assessment were defined as having implemented PATHS with moderate compliance ( $n_{\text{student}} = 801, 49\%$ ); that is, having delivered 67% or more of the lessons in

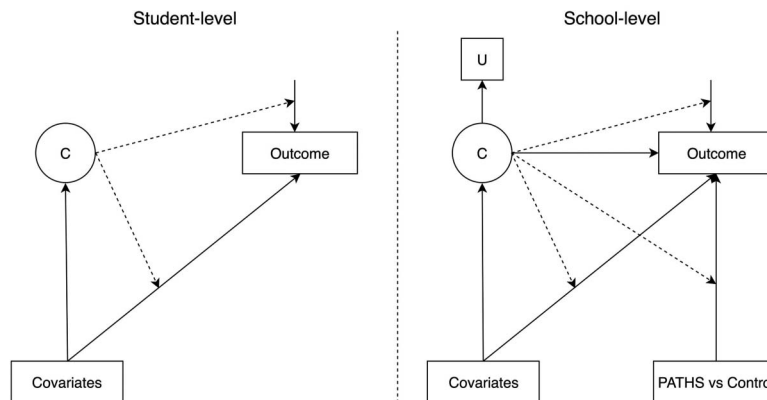


Figure 2. Path diagram for the multilevel complier average causal effect analysis.  $C$  = compliance status (Class 1 = noncompliers; Class 2 = compliers);  $U$  = binary compliance variable for the intervention group (0 = noncompliance; 1 = compliance; -999 = missing for control group).



a year. High compliance, corresponding to delivery of 79% or more of the lessons, was observed in classrooms that fell above the 75th percentile ( $n_{\text{student}} = 408$ , 25%). Our main analysis included models using T1 dosage, given that a combined compliance score from both time points caused computational problems in preliminary analyses. Sensitivity analyses were subsequently conducted using moderate (>36% lessons delivered;  $n_{\text{student}} = 777$ , 29%) and high T2 compliance (>42.8% lessons delivered;  $n_{\text{student}} = 367$ , 13.7%). Descriptive statistics for dosage are presented in Table 1.

## Statistical Analysis

Missing data for the baseline student-level variables ranged between 3.7% and 15.1% and were attributable to student absences (e.g., students changing schools) and parent opt-out. Elevated missingness at follow-up (25.1–25.8%) was attributable to student absences or school drop-out (see Figure 1). Three classrooms at T1 (4.2%) and five at T2 (7.35%) had missing data on dosage as they were not observed due to annual, sick or maternity leave, and lesson observation cancellation. Robust maximum likelihood (MLR) with full information was, therefore, used under the assumption of data missing at random.

Given the complexity associated with data clustering and participant noncompliance, and the fact that joint multilevel and mixture models can be computationally demanding, data in the extant literature are often analyzed dealing only with one of the two issues (Jo, Asparouhov, & Muthén, 2008). Failure, however, to consider compliance, clustering, or both in RCT designs can result to biased results and decreased power to detect ITT and CACE effects (Jo, Asparouhov, & Muthén, 2008; Jo, Asparouhov, Muthén, Ialongo, et al., 2008). Therefore, for each outcome in the current study, two-level ITT and CACE models were fitted in Mplus8.1, with Level 1 representing the students and Level 2 the schools ( $n = 45$  with average cluster size = 115.96).

**ITT analysis.** Treatment assignment (1 = control; 2 = PATHS) along with all student-level and school-level covariates were regressed on the outcome variables for all students in schools that had been randomly allocated in the RCT, irrespective of

whether or not they subsequently participated in the condition to which they were assigned (Gupta, 2011).

**CACE analysis.** For the estimation of CACE we were confident that: (a) because of randomization occurring at the school level, violation of the SUTVA assumption—that is, the interaction among students assigned to different treatment conditions (i.e., contamination)—was improbable (Jo, Asparouhov, Muthén, Ialongo, et al., 2008); (b) treatment assignment was random; and (c) there were neither always-takers nor defiers (i.e., control schools did not have access to PATHS). Although the exclusion restriction assumption might be reasonable under certain conditions (e.g., experiments with blinding), its plausibility in social-behavioral interventions is often questioned (Jo, Asparouhov, Muthén, Ialongo, et al., 2008). In our study, this assumption was less realistic due to the cut-offs used to define compliance. Where this is the case, it is possible to relax the exclusion restriction assuming good predictors of compliance (as demonstrated by Jo, 2002a). However, the effectiveness of such methods are less well known within the multilevel framework (i.e., cluster RCT; Jo, Asparouhov, & Muthén, 2008). Although a simulation study by Jo, Asparouhov, Muthén, Ialongo, et al. (2008) provides some confidence in relaxing this assumption in cluster RCT designs, we decided against it. Missing data within CACE is already complicated (Jo, Ginexi, & Ialongo, 2010) and our model, compared to that of Jo, Asparouhov, Muthén, Ialongo, et al. (2008), was estimated based on a smaller cluster sample ( $n = 45$ ) and missing data on the covariates and outcomes (full information maximum likelihood [FIML]), which in combination with relaxing the exclusion restriction might have further complicated CACE estimation. We therefore relied on the inclusion of good individual-level and school-level predictors, which can mitigate the potential impact of such an assumption being violated (Jo, Asparouhov, Muthén, Ialongo, et al., 2008). Although the choice to comply was at the teacher/classroom level, dosage data were disaggregated to the student level, as information on the class membership for the control schools was not available, thus preventing us from treating classroom as a higher level in our multilevel CACE models.

Given that neither always-takers nor defiers were assumed, the current study focuses only on compliers and never-takers, which we refer to as compliers and noncompliers for simplicity. A multilevel mixture modeling analysis for the CACE models was used as this can accommodate clustered data considering mixture distributions of compliers and noncompliers (Jo, Asparouhov, & Muthén, 2008). CACE was performed using MLR and expectation maximization algorithm (ML-EM; Muthén & Muthén, 1998–2017a), which enables the estimation of the unknown compliance status for the control group (Jo & Muthén, 2001) and in our case, the classrooms that were not observed. Each model was fitted using the moderate and high compliance cut-offs and with double the starting values to ensure that the best log likelihood value was replicated. In instances where a statistically significant intervention effect was observed, an effect size, comparable to Cohen's  $d$  (Cohen, 1992), was calculated using the following formula  $\Delta = \beta/\sigma_e$ , where  $\beta$  represents the treatment beta effect and  $\sigma_e$  indicates the standard deviation of the outcome variable at the student-level (Tymms, 2004).

**Covariates.** Sex (1 = male, 2 = female), student FSM eligibility (0 = no FSM; 1 = FSM), conduct problems, and social-emotional competence were added alongside baseline scores as Level 1 predictors of each outcome and the compliance latent

Table 1  
Means, Standard Deviations, and Internal Consistency  
Coefficient for Study Variables

Outcome variables	<i>M (SD)</i>		$\omega$
	PATHS	Control	
Sex % (male/female)	49.9/50.1	47/53	
Age T1	8.12 (.88)	8.12 (.86)	
Psychological well-being T1	29.76 (4.80)	28.97 (5.04)	.77
Psychological well-being T2	30.16 (4.35)	23.63 (4.97)	.82
Peer social support T1	16.90 (3.34)	16.37 (3.54)	.76
Peer social support T2	17.32 (3.04)	16.94 (3.18)	.80
School connectedness T1	17.75 (2.63)	17.35 (2.88)	.71
School connectedness T2	17.81 (2.45)	17.66 (2.62)	.74
Conduct problems T1	1.08 (1.81)	1.05 (1.80)	.82
Social-emotional competence T1	106.16 (19.95)	103.47 (20.40)	.94
Dosage% T1	66.50 (17.09)	—	—
Dosage% T2	34.98 (16.66)	—	—

Note. T1 = Time; T2 = Time 2.

class. Although it was assessed for the teachers, compliance to an intervention can be influenced by the classroom climate, which has been shown to be affected by student-level characteristics (Domitrovich et al., 2008; Koth, Bradshaw, & Leaf, 2008). For instance, we included children's social-emotional competence as this may be an important factor affecting implementation, particularly for a program focused on SEL (Voegler-Lee, Kupersmidt, Field, & Willoughby, 2012). It is possible that teachers would more likely deliver such an intervention in classes where the perceived need is greater (e.g., with students low in social-emotional competence). School size, poverty (FSM), and minority ethnic composition (EAL) were added as school-level covariates of each outcome but also of compliance latent class, given that previous work has found these to be significant predictors of program implementation (McIntosh, Mercer, Nese, Strickland-Cohen, & Hoselton, 2016; Payne & Eckert, 2010; Payne, Gottfredson, & Gottfredson, 2006). The same covariates were used in ITT models for consistency. Given that, as noted above, classroom was not treated as a level in the analysis, this meant that we could not include teacher level covariates.

## Results

Means, standard deviations, and omega internal consistency coefficients for the study variables are reported in Table 1.

### ITT Analyses

The main ITT analyses, after controlling for student-level and school-level covariates, found a small but statistically significant intervention effect for psychological wellbeing ( $\beta = .81, p < .05; \Delta = .17$ ). However, no such effects were observed for peer social support or school connectedness (see Table 2).

### CACE Analyses

The moderate and high compliance models are reported in Tables 3 and 4, respectively. The entropy, which is a classification quality index (with a range of 0–1), was used to assess the

appropriateness of said models, with higher values indicating that classes are more easily distinguished (Grimm, Ram, & Estabrook, 2017). Classes with no less than 1% total count and high posterior probabilities were also considered acceptable (Jung & Wickrama, 2008). All models met the above criteria and with high entropy values for both moderate (.72–.74) and high compliance (.72–.80). After accounting for baseline scores, conduct problems, social-emotional competence, FSM, sex, school FSM, EAL, and size, models using a moderate compliance cut-off revealed medium to large statistically significant treatment effects for peer social support ( $\beta = 1.98, p < .001; \Delta = .63$ ) and school connectedness ( $\beta = 2.01, p < .001; \Delta = .80$ ). In addition, compared to ITT analyses, the intervention effect size was found to be larger for psychological wellbeing ( $\beta = 2.00, p < .001; \Delta = .43$ ). Intervention effect sizes remained stable in magnitude in the high compliance models ( $\Delta = .43$  for psychological wellbeing;  $\Delta = .63$  for peer social support;  $\Delta = .79$  for school connectedness). Sensitivity analyses using compliance markers derived from T2 dosage data replicated this pattern of findings (see Supplemental Tables S1–S2 of the online supplementary materials). This was unsurprising, as although dosage was reduced by almost half at T2 (see Table 1), only a relatively small percentage of moderate ( $n_{\text{student}} = 166, 20.7\%$ ) and high ( $n_{\text{student}} = 43, 10.5\%$ ) T1 compliers ceased being a complier in the second year of implementation. On the other hand, 32.6% of moderate ( $n_{\text{student}} = 270$ ) and 18% of high ( $n_{\text{student}} = 219$ ) noncompliers at T1 became a complier by T2.

Student-level and school-level covariate effects on compliance are interpreted as logistic and linear regression coefficients, respectively. Social-emotional competence was found to be a statistically significant predictor of moderate compliance in the psychological wellbeing ( $b = -1.59, p < .001$ ; odds ratio [OR] = .21) and peer social support models ( $b = -.96, p < .001$ ; OR = .38). Conduct problems was shown to significantly predict moderate ( $b = -.15, p < .05, OR = .87$ ) and high compliance ( $b = -.16, p < .05; OR = .85$ ) when school connectedness was the outcome. Baseline scores also predicted high compliance in school connectedness ( $b = .86, p < .05; OR = 2.37$ ). It is worth noting that while some of the effects were automatically fixed for two of the high

Table 2  
ITT Effect Sizes for Each Outcome ( $N = 5,218$ )

Predictor	Outcome variables $\beta$ (SE)		
	Psychological wellbeing	Peer social support	School connectedness
<b>Student-level predictors</b>			
Baseline scores	.21 (.02)***	.20 (.03)***	.15 (.03)***
Conduct problems	-.11 (.02)***	-.09 (.02)***	-.16 (.02)***
Social-emotional competence	.04 (.02)	.05 (.02)*	.12 (.02)***
Sex (1 = Female; 2 = Male)	-.00 (.02)	.05 (.02)*	.13 (.02)***
Free school meals (0 = No; 1 = Yes)	-.01 (.02)	-.04 (.02)*	-.06 (.02)**
<b>School-level predictors</b>			
PATHS vs. control	<b>.81 (.32)* <math>\Delta = .17</math></b>	.59 (.31)	.29 (.38)
Free school meals %	.07 (.20)	.50 (.12)***	.27 (.18)
English as additional language %	-.25 (.17)	-.22 (.16)	-.11 (.18)
Size	.19 (.20)	.11 (.15)	.04 (.21)

Note. PATHS = Promoting Alternative Thinking Strategies (PATHS) curriculum. Intent-to-treat effects are in boldface type.

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

Table 3  
CACE Effect Sizes for Each Outcome With Moderate Compliance Levels (N = 5,218)

Predictor	Outcome variables β (SE)					
	Psychological wellbeing		Peer social support		School connectedness	
	Compliers (32%)	Noncompliers (68%)	Compliers (32%)	Noncompliers (68%)	Compliers (31%)	Noncompliers (69%)
<b>Student-level predictors</b>						
Baseline scores	.22 (.05)***	.21 (.02)***	.18 (.03)***	.19 (.03)***	.17 (.04)***	.16 (.03)***
Conduct problems	-.10 (.04)*	-.08 (.03)**	-.09 (.03)**	-.07 (.03)*	-.15 (.04)**	-.11 (.03)***
Social-emotional competence	.09 (.03)**	.06 (.03)*	.02 (.04)	.11 (.03)***	.17 (.04)***	.12 (.02)***
Sex (1 = female; 2 = male)	.02 (.03)	-.01 (.02)	.07 (.04)	.05 (.02)*	.13 (.03)***	.13 (.03)***
Free school meals (0 = no; 1 = yes)	-.02 (.04)	-.01 (.03)	-.00 (.03)	-.04 (.03)	-.08 (.04)***	-.06 (.03)***
<b>School-level predictors</b>						
PATHS vs. control (CACE)	<b>2.00 (.21)*** Δ = .43</b>	—	<b>1.98 (.21)*** Δ = .63</b>	—	<b>2.01 (.21)*** Δ = .80</b>	—
Free school meals %	.01 (.04)	.19 (.27)	.05 (.01)	.41 (.15)**	.01 (.04)	.19 (.27)
English as additional language %	.06 (.04)	.01 (.15)	-.03 (.04)	.10 (.10)	.06 (.04)	.01 (.15)
School size	-.06 (.04)	-.05 (.16)	.03 (.03)	-.09 (.10)	-.06 (.04)	-.05 (.16)

Note. PATHS = Promoting Alternative Thinking Strategies (PATHS) curriculum; CACE = complier average causal effect. CACE effects are in boldface type.  
\* p < .05. \*\* p < .01. \*\*\* p < .001.

compliance models—possibly due to their complexity—high percentages of school EAL appeared to consistently predict low compliance (b = -.53—.58). Results differed somewhat for models using T2 compliance data, where student FSM and school size were shown to be the most common predictors of reduced compliance (see Supplemental Tables S1–S2 in the online supplementary material).

### Discussion

The principal aim of the current study was to advance knowledge and understanding in relation to the impact of an evidence-based SEL intervention on children’s psychological wellbeing, perceptions of peer social support, and school connectedness. This was undertaken in the context of a major randomized trial of the

PATHS curriculum, implemented outside of its country of origin and evaluated independently. As is commonly observed in school-based interventions (Berg et al., 2017; Durlak, 2016; Kam, Greenberg, & Kusché, 2004), implementation in the current study was variable despite successful randomization. Although some previous studies have accounted for implementation effects in schools delivering the PATHS curriculum (Berry et al., 2016; Conduct Problems Prevention Research Group, 1999b; Humphrey, Barlow, & Lendrum, 2018; Humphrey, Hennessey et al., 2018; Kam et al., 2004; Schonfeld et al., 2015), they have failed to include both intervention and control groups of the trial in their analyses. With a few exceptions (noted in the introduction to the current study), the same can be said of the field of school-based intervention research more generally.

Table 4  
CACE Effect Sizes for Each Outcome With High Compliance Levels (N = 5,218)

Predictor	Outcome variables β (SE)					
	Psychological wellbeing		Peer social support		School connectedness	
	Compliers (17%)	Noncompliers (83%)	Compliers (17%)	Noncompliers (83%)	Compliers (17%)	Noncompliers (83%)
<b>Student-level predictors</b>						
Baseline scores	.27 (.08)**	.20 (.03)***	.14 (.04)**	.19 (.03)***	.13 (.05)*	.17 (.03)***
Conduct problems	-.08 (.06)	-.09 (.04)*	-.07 (.07)**	-.08 (.02)**	-.08 (.04)	-.15 (.04)***
Social-emotional competence	.08 (.17)	.07 (.03)*	.04 (.07)	.08 (.03)*	.22 (.06)***	.11 (.03)***
Sex (1 = female; 2 = male)	.05 (.05)	-.01 (.03)	.11 (.06)	.05 (.02)*	.20 (.03)***	.12 (.02)***
Free school meals (0 = no; 1 = yes)	-.03 (.07)	.00 (.04)	-.01 (.05)	-.03 (.03)	-.08 (.04)*	-.06 (.02)*
<b>School-level predictors</b>						
PATHS vs. Control (CACE)	<b>2.00 (.21)*** Δ = .43</b>	—	<b>1.98 (.21)*** Δ = .63</b>	—	<b>2.00 (.21)*** Δ = .79</b>	—
Free school meals %	.07 (.13)	-.18 (.44)	.07 (.32)	.27 (.23)	-.10 (.18)	.25 (.24)
English as additional language %	.04 (.34)	-.17 (2.30)	-.04 (.23)	.10 (.24)	.16 (.13)	-.08 (.23)
School size	-.02 (.09)	.18 (.29)	.01 (.07)	.02 (.37)	-.05 (.08)	.27 (.15)

Note. PATHS = Promoting Alternative Thinking Strategies (PATHS) curriculum; CACE = complier average causal effect. CACE effects are in boldface type.  
\* p < .05. \*\* p < .01. \*\*\* p < .001.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

ITT treatment effects are estimated under the assumption that all participants in the intervention group complied with the intervention (i.e., received the intervention; Jo & Muthén, 2001). However, when noncompliers are also included in the analysis, this can lead to biased and often underestimated intervention effects (Connell, 2009; Jo, Asparouhov, & Muthén, 2008; Jo & Muthén, 2001). Failing to account for compliance variability in the analysis of RCTs can also result in substantially decreased power to detect treatment effects (Jo, Asparouhov, & Muthén, 2008). Therefore, a secondary aim of this study was to demonstrate the estimation of multilevel CACE models for cluster RCT designs, thereby enabling us to determine the extent to which the magnitude of any intervention effects identified in the main ITT analysis changed once compliance was taken into account.

Our results demonstrate that the impact of PATHS on children's psychological wellbeing, peer social support, and school connectedness strongly depends on teachers' intervention compliance. When ITT models were used, an intervention effect was identified only for psychological wellbeing. However, taking compliance into account, we also found intervention effects for peer social support and school connectedness that were statistically significant and medium to large in magnitude. In other words, PATHS was shown to meaningfully improve these outcomes when teachers delivered at least 67% of the curriculum. The effect of PATHS on psychological wellbeing was also stronger in magnitude among teachers demonstrating moderate intervention compliance when compared to that identified through the corresponding ITT analysis. In contrast to other research (Berg et al., 2017; Connell, 2009), identical effect sizes were observed across the moderate and high CACE models, which indicates that further increases in compliance to the PATHS curriculum do not necessarily lead to additional increases in intervention effects. Although our findings align with those by Nagengast et al. (2018), who reported trivial differences between models using different compliance cut-offs (for their text condition outcome), we are cautious in our interpretations given that the high compliance models were estimated based on a smaller sample of compliers (25%,  $n = 408$ ). In addition, this is the first study to estimate CACE effects with FIML within a multilevel framework, the implications of which are largely understudied. It is worth noting that sensitivity analyses using dosage data from the second year of the trial presented a similar picture. No substantial differences were observed between the first and second years of the trial, and effect sizes were mostly identical. This might indicate that, although dosage was generally reduced by T2, the benefits of PATHS were accrued during the first year of the trial, where dosage was higher (though it is also worth noting that most teachers who complied in the initial stages of the trial remained "compliers" throughout the study period).

## Implications

Given that this is the first study to estimate the impact of PATHS on these aspects of children's quality of life, and most importantly through CACE, our findings add new, independent, and rigorous evidence for this program and, by extension, contribute to the growing literature which highlights the utility of SEL as an efficacious means through which to promote socially significant outcomes in school (Durlak et al., 2011; Sklad et al., 2012; Taylor et al., 2017; Wigelsworth et al., 2016). Accordingly, support for

both the underpinning SEL theory (e.g., CASEL, 2003) and PATHS program theory (Greenberg & Kusche, 1993) is implied, though with the important caveat that the magnitude of intervention effects clearly varies as a function of implementation. As a consequence, we propose that intervention logic models should start to reflect this as opposed to simply showing the theorized causal links between program inputs, change mechanisms, and outputs under "ideal" conditions (Humphrey, Lendrum et al., 2016) as in our recent, adapted version of the PATHS logic model (Humphrey, Hennessey et al., 2018). In addition, though it was beyond the scope of the current study given its existing analytical complexity, future research could usefully examine the extent to which the intervention effects observed are mediated by proximal changes in treatment inherent outcomes in a manner consistent with the theorized mechanisms discussed earlier (e.g., is improved psychological wellbeing triggered via enhanced self-awareness and emotional regulation following exposure to PATHS?).

The effects observed in the current study can be considered particularly noteworthy given the evidence of declining levels of wellbeing in the school-aged population in recent years (Children's Society, 2017), the concurrent and prospective associations between wellbeing and academic attainment (Gutman & Vorhaus, 2012), and the established longer-term links between improved wellbeing and lower mortality rates (Chida & Steptoe, 2008). Similarly, the peer relations and support effects are promising given the primacy of positive peer relations in child and adolescent development (Bagwell & Schmidt, 2011; Bierman, 2005; Kendrick et al., 2012), in addition to their obvious utility in the educational context. We know that students are more resilient and responsive academically in classrooms and schools in which they feel safe and cared about, and where learning and relationship building are facilitated (Durlak et al., 2011; Humphrey, 2013; Kochel et al., 2017). In the same way, students having a sense of school belonging are less likely to report poor mental health, social rejection, and school problems, such as school disengagement (Anderman, 2002; Kia-Keating & Ellis, 2007). These findings also highlight the utility of CACE in determining what could be regarded as the "true potential" of interventions such as PATHS (e.g., the actual effects of the intervention itself as opposed to the effects of being randomized to receive the intervention, as in ITT; Peugh et al., 2017). In our case, we find that exposure to at least 67% of the lessons during the first year of the trial optimized the impact of the intervention. However, this comes with a caveat, in that ITT analyses may still be considered more indicative of the likely intervention effects when interventions are implemented outside of trial settings, as we know that noncompliance is highly probable, and indeed may increase significantly, out in the "real-world" of schools (Humphrey, Lendrum et al., 2016).

Our CACE findings demonstrate that, beyond a certain point, more is not necessarily better (Durlak & DuPre, 2008). Given, however, the low mean compliance (relative to developer expectations) in the current and other studies (Berry et al., 2016; Conduct Problems Prevention Research Group, 1999b), they also raise the question of whether a shorter version of PATHS would be as or more effective. Although we are unable to currently provide a definitive answer, it is worth bearing in mind that had they followed an abridged delivery model, teachers would likely have to comply 100% with the intervention to replicate the CACE findings reported here. The cost-benefit effectiveness of such an approach

could, however, be improved if the levels of implementation necessary to achieve the best results are known (Durlak & DuPre, 2008). Therefore, future research should consider establishing different implementation thresholds for different outcomes.

Given the primacy of implementation variability in our findings, a natural corollary is for us to consider theory and research on what drives this in school-based interventions. Theoretical models (e.g., Domitrovich et al., 2008), reviews of research (e.g., Durlak & DuPre, 2008), and individual studies (Domitrovich et al., 2015) highlight a range of factors operating at different levels. These include preplanning and foundations (e.g., the need, readiness and capacity for change in the school), the implementation support system (e.g., the provision and form of initial training and ongoing coaching or other support), the implementation environment (e.g., compositional and contextual characteristics, leadership support), implementer factors (e.g., the “will and skill” of teachers), and the characteristics of the intervention itself (e.g., its complexity and prescriptiveness). In terms of the current study, qualitative IPE data revealed solid foundations, with clearly identified student needs and desire for change among participating schools (Humphrey, Hennessey et al., 2018). Furthermore, as noted earlier, the implementation support system was strong, with teachers in PATHS schools receiving initial and follow-up training in addition to ongoing technical support and assistance from trained coaches. However, in terms of the implementation environment, our IPE data revealed issues of competing priorities, curriculum demands and lack of time. In addition, teacher attitudes toward PATHS varied, with opinion divided, in particular, regarding the structured, prescriptive nature of the intervention (Humphrey, Hennessey et al., 2018).

Unlike others (Voegler-Lee et al., 2012) we found that the odds of being a complier increased in classrooms with students with lower average social-emotional competence scores. This may be a reflection of higher perception of need in such settings (a proposition given some support by qualitative data in the aforementioned IPE; Humphrey, Hennessey et al., 2018) and aligns with Kowalski, Pretti-Frontczak, and Johnson (2009), who found that teachers feel that social-emotional skills are more important than academic skills. Inconsistent with existing literature (Payne et al., 2006), we also found that teachers were less likely to comply in schools with higher percentages of students speaking EAL. This highlights how natural variation in school compositional and contextual characteristics may underpin between-school differences in implementation and may reflect the importance of appropriate adaptation for ethnic minority populations (Durlak & DuPre, 2008). It is possible, for instance, that teachers felt they were unable to adapt PATHS to the needs and preferences of children whose first language was not English (Humphrey, Hennessey et al., 2018). Finally, conduct problems also influenced compliance, but only in one of the models. This might be explained by the low percentage of students considered as exhibiting high levels of conduct problems ( $n = 482, 9.2\%$ ). It is also likely that such difficulties are less salient for teachers delivering SEL interventions, especially compared to social-emotional competence. It is worth noting that although lower scores of social-emotional competence were a key driver of increased compliance in the first year of the trial, this was not the case for the following year. At T2, poverty and school size were shown to be influencing compliance,

indicating that different factors may influence implementation over time.

## Strengths and Limitations

The current study has numerous strengths that increase confidence in the security of the findings reported herein. First, we used a large and robust cluster RCT design with appropriate analyses that took both compliance and data clustering into account, and ours is among the few available studies that provide a multilevel CACE model, and the first, to our knowledge, to accommodate missing data within this framework. The use of cluster-randomization minimized the possibility of diffusion and or contamination, and the randomization process itself was conducted independently of the research team. Trial school composition mirrored that of primary schools in England with the exception of EAL and FSM. Good predictors of compliance were used, and the outcome measures were psychometrically robust and theoretically plausible, but not intervention specific (e.g., not “inherent to treatment”). Finally, the application of CACE allowed for an unbiased estimate of the effects of PATHS in which intervention compliance was taken into account.

There are, however, a few limitations that need to be borne in mind. First, although CACE estimation requires a binary compliance indicator, in the case of school-based interventions, partial compliance is often observed, meaning that there is potential for the exclusion restriction assumption to be violated (Education Endowment Foundation, 2018). Put more simply, CACE assumes that those deemed to be noncompliers in the intervention arm accrue no benefit from the intervention, when in fact this may not be the case—it really depends on how or where the compliance threshold is set. In the current study and others like it (e.g., Berg et al., 2017), this issue was partially addressed via sensitivity analyses in which the compliance threshold was manipulated in order to determine whether this made a substantive difference. To make our findings more robust, we relied on the inclusion of good predictors of compliance. However, this is often not ideal given that bias due to the violation of the exclusion restriction assumption cannot be completely eliminated (Jo, 2002b), and so alternative approaches should be considered. This might include prespecifying minimal and optimal compliance thresholds or using related instrumental variable approaches that allow for continuous measures of compliance (Education Endowment Foundation, 2018; Gerber & Green, 2012).

Second, because CACE requires a single indicator, only intervention dosage data were used in our analysis. This means that other potentially important implementation dimensions (e.g., procedural fidelity) were neglected. Alternative approaches, such as using multiple implementation dimensions to create a composite compliance indicator, or estimating causal effects separately for each dimension, were considered but dismissed. In the case of the former, it would have been challenging to combine dimensions in a meaningful way given the lack of theoretical or empirical benchmarks. In the case of the latter, interpretation would have been problematic as each CACE estimate would potentially be defined by a different sample of compliers (e.g., those compliant in terms of dosage may not be compliant in terms of procedural fidelity, and so on).

Third, the inherent limitations of the implementation measure itself must be acknowledged. Chief among these is that it provided only a relative “snapshot” of implementation. Clearly, repeated observations throughout the trial would have been preferable, but would have produced unacceptable levels of data burden for participating schools. We also note that some major observational studies of temporal patterns in implementation have evidenced considerable stability (e.g., Hansen et al., 2013), including some that have focused on PATHS. For example, Domitrovich, Gest, Jones, Gill, and Sanford DeRousie (2010) found no significant changes in PATHS dosage, fidelity or participant responsiveness in implementation data collected monthly over the course of a school year. We were confident in the accuracy of the data generated, as independent observations have been found to be more robust than teacher self-ratings (Hansen, 2014). Although observations were scheduled with teachers in advance, meaning that in theory they could simply teach the scheduled lesson while being observed, our data demonstrated that this was clearly not the case. For example, projected dosage was as low as 24% for one class in the first year of the trial.

Fourth, it is worth noting the potential limitations arising from our use of solely self-reported outcome measures. There have been substantial discrepancies reported between parent- and child-reported quality of life (Jozefiak, Larsson, Wichstrøm, Mattejat, & Ravens-Sieberer, 2008; Upton, Lawford, & Eiser, 2008) and indeed this is the case for the psychological wellbeing and peer social support domains of KS27 (Robitail, Simeoni et al., 2007). For the Kidscreen measure specifically, discrepancies have been shown to be the result of differences in the response style and reasoning, and item interpretation (Davis et al., 2007). It has, therefore, been argued that comparing the ratings between self- and proxy-report quality of life might be meaningless and potentially misleading (Bagheri, Jafari, Tashakor, Kouhpayeh, & Riazi, 2014). Although some suggest that both sources should be used, whenever possible (Erhart, Ellert, Kurth, & Ravens-Sieberer, 2009), others argue that parent-report might not be appropriate in capturing the construct of quality of life, which by definition concerns the individual’s subjective perspective (Jozefiak, 2014; Wallander & Koot, 2016). Therefore, self-report should be the primary source of information (Wallander & Koot, 2016), especially in healthy populations (Rajmil, López, López-Aguilà, & Alonso, 2013).

However, with specific reference to perceptions of peer relations and support, it has been argued that reliance on solely self-report measures fails to fully capture children’s actual social interactions (Parker & Asher, 1993). For instance, peer nomination is one of the most widely used methods of sociometric status (see Gifford-Smith & Brownell, 2003; Pepler & Craig, 2003, for a detailed comparison of methods). However, peer-rated assessments can be influenced by reputational bias (Hymel, Wagner, & Butler, 1990) and were shown to correlate poorly with self-ratings on specific aspects of social relationships, such as bullying and victimization (Branson & Cornell, 2009). Similarly, although in theory teachers could be reliable observers of children’s social networks (Gest, 2006), their ratings of children’s friendships have recently been shown to be less accurate than self-report (Meyer & Ostrosky, 2018). Given that friendships and peer relationships are neither group- nor individual-specific constructs, their measurement poses a challenge (Gifford-Smith & Brownell, 2003). Therefore, future

research should carefully consider such measurement issues and, if possible, include multiple (e.g., through latent modeling as in Nangle, Erdley, Newman, Mason, & Carpenter, 2010) sources of information that fit with the research questions being examined.

Fifth, although relevant student and school characteristics were used as predictors of compliance, we were unable to model teacher characteristics, because of the lack of classroom membership information for the control schools. Subsequently, it was necessary to disaggregate the compliance data to the student level, a method that was shown to perform well within CACE analysis (Jo, Asprouhov, Muthén, Ialongo, et al., 2008). However, there was a missed opportunity to specifically account for the classroom-level effects by exploring a more comprehensive model through a three-level CACE with students in Level 1, teachers in Level 2, and schools in Level 3. Future studies should therefore seek to address this limitation. Finally, because of computational issues, we were unable to test the combined effects of compliance from both years of the trial. However, results from sensitivity analyses provided support for invariant effects between models using dosage data from different time points.

## Recommendations for Future Research

Much of the current evidence on the efficacy of SEL interventions relies on the ITT model, and those exploring implementation effects routinely fail to include both intervention and control groups in their analyses. Although ITT effects offer important information, especially for understanding whether school-based interventions can work under real world conditions (Berg et al., 2017), their effects might be underestimated (Connell, 2009). Estimating the effect of compliance to an intervention can provide more robust and generalizable estimates, which can be useful in policy-making (Stuart et al., 2008). Drawing from the current study, we therefore make several recommendations for future research. We agree with CACE advocates (e.g., Peugh et al., 2017) that its use should become more widespread in school psychology. It is important, however, to retain ITT analysis; one should not replace the other. As described above, ITT alongside CACE can certainly provide more robust and useful information about intervention effects than ITT alone (Stuart et al., 2008). However, routine incorporation of CACE estimation will likely require agreement between program developers and researchers on the definition of compliance in the context of a given intervention so that this can be specified a priori (Education Endowment Foundation, 2018; Schochet & Chiang, 2011). Such a definition should be derived from intervention theory, and the notion of critical component analysis, wherein the active ingredients of an intervention are theorized and tested, may be useful here. Blase and Fixsen (2013) argued that proposed core intervention components can be operationalized as contextual factors (e.g., the particular setting in which the intervention occurs), structural elements (e.g., the obligatory number and/or sequence of sessions) and/or specific practices (e.g., reinforcing intervention concepts and ideas outside of sessions to promote generalization of learning); a combination of these could feasibly be used to generate a meaningful compliance indicator for a given intervention. However, this issue requires further attention. Intervention manuals rarely specify (proposed) critical components, and indeed, (Durlak, 2010) noted that, “the core elements of most psychosocial and educational interven-

tions are unknown” (p. 351). Therefore, where this is the case, we suggest that different compliance thresholds are considered through sensitivity analysis as in the current and other studies (Berg et al., 2017; Sagarin et al., 2014).

Following others (Nagengast et al., 2018; Stuart et al., 2008) we advise future researchers to collect data that are expected to affect compliance rates and that are in line with the focus of their intervention. For instance, in the current study, social-emotional competence—which is directly related to the focus of SEL interventions—was shown to be a significant predictor of compliance. The inclusion of strong predictors also decreases sensitivity to violation of underlying assumptions such as the exclusion restriction and increases identifiability of CACE effects when such assumptions are relaxed (Jo, Asparouhov, Muthén, Ialongo, et al., 2008). Although techniques that relax the exclusion restriction have received increased support over the years, their applicability in multilevel CACE is well less studied. In addition, much of the current work relies on listwise deletion (Berg et al., 2017; Jo, 2002b; Jo, Asparouhov, Muthén, Ialongo, et al., 2008) despite repeated recommendations against it (Harel, Zimmerman, & Dekhtyar, 2008) and the possibility of accommodating missing data within the ML-EM framework (Jo, 2002b). Therefore, further work is needed to explore the plausibility of real-life scenarios, such as how to work with cluster RCT data with noncompliance, a smaller number of clusters, and missing data on the covariates and outcomes, the latter of which is commonly observed in school-based interventions (see Taylor et al., 2017).

Finally, compliance in school trials is often decided at the cluster level (e.g., teacher or school) requiring disaggregation of data to the student level. Although a simulation study by Jo, Asparouhov, Muthén, Ialongo, et al. (2008) has found no issues with modeling individual-level compliance variables with low within-level variation (i.e., with high intra cluster correlation coefficient [ICC]), we found that a combined dosage variable from both time points of the trial with ICC = 1 was causing computational problems in the current study. Given that the impact of such variables on the quality of the estimation is unclear (Muthén & Muthén, 1998–2017b), further investigation is in order. Until such time, we advise that when these are used, their impact on model estimation is considered.

In conclusion, considering both ITT and CACE analyses, the current study has rigorously demonstrated that the PATHS curriculum significantly improves children’s psychological wellbeing, peer social support, and school connectedness. From a methodological perspective, the current study also highlights the utility of CACE estimation for school-based RCTs in school psychology, as our analyses indicated that the magnitude of these effects is increased when intervention compliance is taken into account. However, further efforts to optimize implementation are required if the true potential of PATHS (and by extension, other SEL interventions) is to be realized.

## References

Anderman, E. M. (2002). School effects on psychological outcomes during adolescence. *Journal of Educational Psychology, 94*, 795–809. <http://dx.doi.org/10.1037/0022-0663.94.4.795>

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American*

*Statistical Association, 91*, 444–455. <http://dx.doi.org/10.1080/01621459.1996.10476902>

Ashdown, D. M., & Bernard, M. E. (2012). Can explicit instruction in social and emotional learning skills benefit the social-emotional development, well-being, and academic achievement of young children? *Early Childhood Education Journal, 39*, 397–405. <http://dx.doi.org/10.1007/s10643-011-0481-x>

Bagheri, Z., Jafari, P., Tashakor, E., Kouhpayeh, A., & Riazi, H. (2014). Assessing whether measurement invariance of the KIDSCREEN-27 across child-parent dyad depends on the child gender: A multiple group confirmatory factor analysis. *Global Journal of Health Science, 6*, 142–153. <http://dx.doi.org/10.5539/gjhs.v6n5p142>

Bagwell, C. L., & Schmidt, M. E. (2011). *Friendships in childhood and adolescence*. New York, NY: Guilford Press.

Berg, J. K., Bradshaw, C. P., Jo, B., & Ialongo, N. S. (2017). Using complier average causal effect estimation to determine the impacts of the Good Behavior Game preventive intervention on teacher implementers. *Administration and Policy in Mental Health and Mental Health Services Research, 44*, 558–571. <http://dx.doi.org/10.1007/s10488-016-0738-1>

Berry, V., Axford, N., Blower, S., Taylor, R. S., Edwards, R. T., Tobin, K., . . . Bywater, T. (2016). The effectiveness and micro-costing analysis of a universal, school-based, social-emotional learning programme in the U. K.: A cluster-randomised controlled trial. *School Mental Health, 8*, 238–256. <http://dx.doi.org/10.1007/s12310-015-9160-1>

Bierman, K. L. (2005). *Peer rejection: Developmental processes and intervention strategies*. New York, NY: Guilford Press.

Blase, K., & Fixsen, D. (2013). *Core intervention components: Identifying and operationalizing what makes programs work* (ASPE Research Brief). Retrieved from [https://aspe.hhs.gov/system/files/pdf/180286/rb\\_CoreIntervention.pdf](https://aspe.hhs.gov/system/files/pdf/180286/rb_CoreIntervention.pdf)

Bond, L., Butler, H., Thomas, L., Carlin, J., Glover, S., Bowes, G., & Patton, G. (2007). Social and school connectedness in early secondary school as predictors of late teenage substance use, mental health, and academic outcomes. *Journal of Adolescent Health, 40*, 357.e9–357.e18. <http://dx.doi.org/10.1016/j.jadohealth.2006.10.013>

Bonell, C., Humphrey, N., Fletcher, A., Moore, L., Anderson, R., & Campbell, R. (2014). Why schools should promote students’ health and wellbeing. *British Medical Journal, 348*, g3078. <http://dx.doi.org/10.1136/bmj.g3078>

Brackett, M. A., & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin, 29*, 1147–1158. <http://dx.doi.org/10.1177/0146167203254596>

Branson, C. E., & Cornell, D. G. (2009). A comparison of self and peer reports in the assessment of middle school bullying. *Journal of Applied School Psychology, 25*, 5–27. <http://dx.doi.org/10.1080/15377900802484133>

Bywater, T., & Sharples, J. (2012). Effective evidence-based interventions for emotional well-being: Lessons for policy and practice. *Research Papers in Education, 27*, 389–408. <http://dx.doi.org/10.1080/02671522.2012.690242>

Catalano, R. F., Haggerty, K. P., Oesterle, S., Fleming, C. B., & Hawkins, J. D. (2004). The importance of bonding to school for healthy development: Findings from the Social Development Research Group. *The Journal of School Health, 74*, 252–261. <http://dx.doi.org/10.1111/j.1746-1561.2004.tb08281.x>

Center for the Study and Prevention of Violence. (2012). *Promoting alternative thinking strategies*. Retrieved from <https://www.blueprintsprograms.org/factsheet/promoting-alternative-thinking-strategies-paths>

Chida, Y., & Steptoe, A. (2008). Positive psychological well-being and mortality: A quantitative review of prospective observational studies. *Psychosomatic Medicine, 70*, 741–756. <http://dx.doi.org/10.1097/PSY.0b013e31818105ba>

- Children's Society. (2017). *The good childhood report 2017*. Retrieved from [https://www.childrenssociety.org.uk/sites/default/files/the-good-childhood-report-2017\\_full-report\\_0.pdf](https://www.childrenssociety.org.uk/sites/default/files/the-good-childhood-report-2017_full-report_0.pdf)
- Cogo-Moreira, H., Brandão de Ávila, C. R., Ploubidis, G. B., & Mari, J. J. (2013). Effectiveness of music education for the improvement of reading skills and academic achievement in young poor readers: A pragmatic cluster-randomized, controlled clinical trial. *PLoS ONE*, *8*, e59984. <http://dx.doi.org/10.1371/journal.pone.0059984>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. <http://dx.doi.org/10.1037/0033-2909.112.1.155>
- Collaborative for Academic, Social, and Emotional Learning. (2003). *Safe and sound: An educational leader's guide to evidence-based Social and Emotional Learning (SEL) programs*. Chicago, IL: Author.
- Conduct Problems Prevention Research Group. (1999a). Initial impact of the Fast Track prevention trial for conduct problems: I. The high-risk sample. *Journal of Consulting and Clinical Psychology*, *67*, 631–647. <http://dx.doi.org/10.1037/0022-006X.67.5.631>
- Conduct Problems Prevention Research Group. (1999b). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology*, *67*, 648–657. <http://dx.doi.org/10.1037/0022-006X.67.5.648>
- Connell, A. M. (2009). Employing complier average causal effect analytic methods to examine effects of randomized encouragement trials. *The American Journal of Drug and Alcohol Abuse*, *35*, 253–259. <http://dx.doi.org/10.1080/00952990903005882>
- Connell, A. M., Dishion, T. J., Yasui, M., & Kavanagh, K. (2007). An adaptive approach to family intervention: Linking engagement in family-centered intervention to reductions in adolescent problem behavior. *Journal of Consulting and Clinical Psychology*, *75*, 568–579. <http://dx.doi.org/10.1037/0022-006X.75.4.568>
- Corcoran, R. P., Cheung, A., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review*, *25*, 56–72. <http://dx.doi.org/10.1016/j.edurev.2017.12.001>
- Crosnoe, R., Cavanagh, S., & Elder, G. H., Jr. (2003). Adolescent friendships as academic resources: The intersection of friendship, race, and school disadvantage. *Sociological Perspectives*, *46*, 331–352. <http://dx.doi.org/10.1525/sop.2003.46.3.331>
- Davis, E., Nicolas, C., Waters, E., Cook, K., Gibbs, L., Gosch, A., & Ravens-Sieberer, U. (2007). Parent-proxy and child self-reported health-related quality of life: Using qualitative methods to explain the discordance. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*, *16*, 863–871. <http://dx.doi.org/10.1007/s11136-007-9187-3>
- Denham, S. A. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development*, *17*, 57–89. [http://dx.doi.org/10.1207/s15566935eed1701\\_4](http://dx.doi.org/10.1207/s15566935eed1701_4)
- Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach-Major, S., & Queenan, P. (2003). Preschool emotional competence: Pathway to social competence? *Child Development*, *74*, 238–256. <http://dx.doi.org/10.1111/1467-8624.00533>
- Department for Education. (2012, January). *Schools, pupils, and their characteristics* (SFR10/2012). Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/219260/sfr10-2012.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/219260/sfr10-2012.pdf)
- Department for Education. (2013, January). *Special educational needs in England*. (SFR 30/2013). Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/225699/SFR30-2013\\_Text.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/225699/SFR30-2013_Text.pdf)
- Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., . . . Ialongo, N. S. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion*, *1*, 6–28. <http://dx.doi.org/10.1080/1754730X.2008.9715730>
- Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *The Journal of Primary Prevention*, *28*, 67–91. <http://dx.doi.org/10.1007/s10935-007-0081-0>
- Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., & Sanford Derosie, R. M. (2010). Implementation quality: Lessons learned in the context of the Head Start REDI trial. *Early Childhood Research Quarterly*, *25*, 284–298. <http://dx.doi.org/10.1016/j.ecresq.2010.04.001>
- Domitrovich, C. E., Pas, E. T., Bradshaw, C. P., Becker, K. D., Keperling, J. P., Embry, D. D., & Ialongo, N. (2015). Individual and school organizational factors that influence implementation of the PAX Good Behavior Game intervention. *Prevention Science*, *16*, 1064–1074. <http://dx.doi.org/10.1007/s11121-015-0557-8>
- Durlak, J. (1997). *Successful prevention programs for children and adolescents*. New York, NY: Plenum Press. <http://dx.doi.org/10.1007/978-1-4899-0065-4>
- Durlak, J. (2010). The importance of doing well in whatever you do: A commentary on the special ed., "Implementation research in early childhood education." *Early Childhood Research Quarterly*, *25*, 348–357. <http://dx.doi.org/10.1016/j.ecresq.2010.03.003>
- Durlak, J. A. (2016). Programme implementation in social and emotional learning: Basic issues and research findings. *Cambridge Journal of Education*, *46*, 333–345. <http://dx.doi.org/10.1080/0305764X.2016.1142504>
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327–350. <http://dx.doi.org/10.1007/s10464-008-9165-0>
- 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>
- Education Endowment Foundation. (2018). *Statistical analysis guidance for EEF evaluations*. Retrieved from [https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing\\_a\\_Protocol\\_or\\_SAP/EEF\\_statistical\\_analysis\\_guidance\\_2018.pdf](https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf)
- Erhart, M., Ellert, U., Kurth, B. M., & Ravens-Sieberer, U. (2009). Measuring adolescents' HRQoL via self reports and parent proxy reports: An evaluation of the psychometric properties of both versions of the KINDL-R instrument. *Health and Quality of Life Outcomes*, *7*, 77. <http://dx.doi.org/10.1186/1477-7525-7-77>
- Forman, S. G. (2015). *Implementation of mental health programs in schools: A change agent's guide*. Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/14597-000>
- Fosco, G. M., Van Ryzin, M. J., Connell, A. M., & Stormshak, E. A. (2016). Preventing adolescent depression with the family check-up: Examining family conflict as a mechanism of change. *Journal of Family Psychology*, *30*, 82–92. <http://dx.doi.org/10.1037/fam0000147>
- Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis and interpretation*. New York, NY: Norton.
- Gest, S. D. (2006). Teacher reports of children's friendships and social groups: Agreement with peer reports and implications for studying peer similarity. *Social Development*, *15*, 248–259. <http://dx.doi.org/10.1111/j.1467-9507.2006.00339.x>
- Gifford-Smith, M. E., & Brownell, C. A. (2003). Childhood peer relationships: Social acceptance, friendships, and peer networks. *Journal of School Psychology*, *41*, 235–284. [http://dx.doi.org/10.1016/S0022-4405\(03\)00048-7](http://dx.doi.org/10.1016/S0022-4405(03)00048-7)
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology & Psychiatry & Allied*



- Disciplines*, 38, 581–586. <http://dx.doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Greenberg, M. T., & Kusche, C. A. (1993). *Promoting social and emotional development in deaf children: The PATHS project*. Seattle: University of Washington Press.
- Greenberg, M. T., Kusche, C. A., Cook, E. T., & Quamma, J. P. (1995). Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. *Development and Psychopathology*, 7, 117–136. <http://dx.doi.org/10.1017/S0954579400006374>
- Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System (SSIS) rating scales manual*. Minneapolis, MN: Pearson.
- Gresham, F. M., Elliott, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of the Social Skills Rating System to the Social Skills Improvement System: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly*, 26, 27–44. <http://dx.doi.org/10.1037/a0022662>
- Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling: Structural equation and multilevel modeling approaches*. New York, NY: Guilford Press.
- Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, 2, 109–112. <http://dx.doi.org/10.4103/2229-3485.83221>
- Gutman, L. M., & Vorhaus, J. (2012). *The impact of pupil behaviour and wellbeing on educational outcomes* (DFE-RR253). Retrieved from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/219638/DFE-RR253.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/219638/DFE-RR253.pdf)
- Hansen, W. (2014). Measuring fidelity. In Z. Sloboda & H. Petras (Eds.), *Defining prevention science* (pp. 335–359). New York, NY: Springer. [http://dx.doi.org/10.1007/978-1-4899-7424-2\\_15](http://dx.doi.org/10.1007/978-1-4899-7424-2_15)
- Hansen, W. B., Pankratz, M. M., Dusenbury, L., Giles, S. M., Bishop, D. C., Albritton, J., . . . Strack, J. (2013). Styles of adaptation: The impact of frequency and valence of adaptation on preventing substance use. *Health Education*, 113, 345–363. <http://dx.doi.org/10.1108/09654281311329268>
- Harel, O., Zimmerman, R., & Dekhtyar, O. (2008). Approaches to the handling of missing data in communication research. In A. Hayes, M. Slater, & L. Snyder (Eds.), *The SAGE sourcebook of advanced data analysis methods for communication research* (pp. 349–372). Thousand Oaks, CA : Sage Publications, Inc. <http://dx.doi.org/10.4135/9781452272054.n12>
- Hay, D. F., Payne, A., & Chadwick, A. (2004). Peer relations in childhood. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 45, 84–108. <http://dx.doi.org/10.1046/j.0021-9630.2003.00308.x>
- Henry, K. L., & Slater, M. D. (2007). The contextual effect of school attachment on young adolescents' alcohol use. *The Journal of School Health*, 77, 67–74. <http://dx.doi.org/10.1111/j.1746-1561.2007.00169.x>
- Humphrey, N. (2013). *Social and emotional learning: A critical appraisal*. London, UK: SAGE Publications. <http://dx.doi.org/10.4135/9781446288603>
- Humphrey, N., Barlow, A., & Lendrum, A. (2018). Quality matters: Implementation moderates student outcomes in the PATHS curriculum. *Prevention Science*, 19, 197–208. <http://dx.doi.org/10.1007/s11121-017-0802-4>
- Humphrey, N., Barlow, A., Wigelsworth, M., Lendrum, A., Pert, K., Joyce, C., . . . Turner, A. (2015). *Promoting Alternative Thinking Strategies (PATHS): Evaluation report and executive summary*. Retrieved from [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/EEF\\_Project\\_Report\\_PromotingAlternativeThinkingStrategies.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_PromotingAlternativeThinkingStrategies.pdf)
- Humphrey, N., Barlow, A., Wigelsworth, M., Lendrum, A., Pert, K., Joyce, C., . . . Turner, A. (2016). A cluster randomized controlled trial of the Promoting Alternative Thinking Strategies (PATHS) curriculum. *Journal of School Psychology*, 58, 73–89. <http://dx.doi.org/10.1016/j.jsp.2016.07.002>
- Humphrey, N., Hennessey, A., Lendrum, A., Wigelsworth, M., Turner, A., Panayiotou, M., . . . Calam, R. (2018). The PATHS curriculum for promoting social and emotional well-being among children aged 7–9 years: A cluster RCT. *Public Health Research*, 6, 1–116. <http://dx.doi.org/10.3310/phr06100>
- Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). *Implementation and process evaluation (IPE) for interventions in educational settings: A synthesis of the literature*. Retrieved from [https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting\\_up\\_an\\_Evaluation/IPE\\_Review\\_Final.pdf](https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Review_Final.pdf)
- Hymel, S., Wagner, E., & Butler, L. J. (1990). Reputational bias: View from the peer group. In S. R. Asher & J. D. Coie (Eds.), *Peer rejection in childhood* (pp. 156–188). New York, NY: Cambridge University Press.
- Imuta, K., Henry, J. D. b., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. *Developmental Psychology*, 52, 1192–1205. <http://dx.doi.org/10.1037/dev0000140>
- Jago, R., Edwards, M. J., Sebire, S. J., Tomkinson, K., Bird, E. L., Banfield, K., . . . Blair, P. S. (2015). Effect and cost of an after-school dance programme on the physical activity of 11–12 year old girls: The Bristol Girls Dance Project, a school-based cluster randomised controlled trial. *The International Journal of Behavioral Nutrition and Physical Activity*, 12, 128. <http://dx.doi.org/10.1186/s12966-015-0289-y>
- Jimerson, S. R., Campos, E., & Greif, J. L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *California School Psychologist*, 8, 7–27. <http://dx.doi.org/10.1007/BF03340893>
- Jo, B. (2002a). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics*, 27, 385–409. <http://dx.doi.org/10.3102/10769986027004385>
- Jo, B. (2002b). Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance. *Statistics in Medicine*, 21, 3161–3181. <http://dx.doi.org/10.1002/sim.1267>
- Jo, B., Asparouhov, T., & Muthén, B. O. (2008). Intention-to-treat analysis in cluster randomized trials with noncompliance. *Statistics in Medicine*, 27, 5565–5577. <http://dx.doi.org/10.1002/sim.3370>
- Jo, B., Asparouhov, T., Muthén, B. O., Ialongo, N. S., & Brown, C. H. (2008). Cluster randomized trials with treatment noncompliance. *Psychological Methods*, 13, 1–18. <http://dx.doi.org/10.1037/1082-989X.13.1.1>
- Jo, B., Ginexi, E. M., & Ialongo, N. S. (2010). Handling missing data in randomized experiments with noncompliance. *Prevention Science*, 11, 384–396. <http://dx.doi.org/10.1007/s11121-010-0175-4>
- Jo, B., & Muthén, B. O. (2001). Modeling of intervention effects with noncompliance: A latent variable approach for randomized trials. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in Structural Equation Modeling* (pp. 57–87). Mahwah, NJ: Erlbaum, Inc.
- Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*, 105, 2283–2290. <http://dx.doi.org/10.2105/AJPH.2015.302630>
- Jozefiak, T. (2014). Can we trust in parents' report about their children's well-being? In A. Ben-Arieh, F. Casas, I. Frönes, & J. E. Korbin (Eds.), *Handbook of child well-being: Theories, methods and policies in global perspective* (pp. 577–578). Dortmund, the Netherlands: Springer.
- Jozefiak, T., Larsson, B., Wichstrøm, L., Matthejat, F., & Ravens-Sieberer, U. (2008). Quality of life as reported by school children and their parents: A cross-sectional survey. *Health and Quality of Life Outcomes*, 6, 34. <http://dx.doi.org/10.1186/1477-7525-6-34>
- Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality*

- Psychology Compass*, 2, 302–317. <http://dx.doi.org/10.1111/j.1751-9004.2007.00054.x>
- Kam, C.-M., Greenberg, M. T., & Kusché, C. A. (2004). Sustained effects of the PATHS curriculum on the social and psychological adjustment of children in special education. *Journal of Emotional and Behavioral Disorders*, 12, 66–78. <http://dx.doi.org/10.1177/10634266040120020101>
- Kendrick, K., Jutengren, G., & Stattin, H. (2012). The protective role of supportive friends against bullying perpetration and victimization. *Journal of Adolescence*, 35, 1069–1080. <http://dx.doi.org/10.1016/j.adolescence.2012.02.014>
- Keyes, C. L. M. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology*, 73, 539–548. <http://dx.doi.org/10.1037/0022-006X.73.3.539>
- Keyes, C. L. M. (2006). The subjective well-being of America's youth: Toward a comprehensive assessment. *Adolescent & Family Health*, 4, 3–11. Retrieved from <https://www.researchgate.net/publication/232509662>
- Kia-Keating, M., & Ellis, B. H. (2007). Belonging and connection to school in resettlement: Young refugees, school belonging, and psychosocial adjustment. *Clinical Child Psychology and Psychiatry*, 12, 29–43. <http://dx.doi.org/10.1177/1359104507071052>
- Kochel, K. P., Bagwell, C. L., Ladd, G. W., & Rudolph, K. D. (2017). Do positive peer relations mitigate transactions between depressive symptoms and peer victimization in adolescence? *Journal of Applied Developmental Psychology*, 51, 44–54. <http://dx.doi.org/10.1016/j.appdev.2017.04.003>
- Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2008). A multilevel study of predictors of student perceptions of school climate: The effect of classroom-level factors. *Journal of Educational Psychology*, 100, 96–104. <http://dx.doi.org/10.1037/0022-0663.100.1.96>
- Kowalski, K., Pretti-Frontczak, K., & Johnson, L. (2009). Preschool teachers' beliefs concerning the importance of various developmental skills and abilities. *Journal of Research in Childhood Education*, 16, 5–14. <http://dx.doi.org/10.1080/02568540109594970>
- Kusche, C., & Greenberg, M. T. (1994). *The PATHS curriculum*. Seattle, WA: Developmental Research and Programs.
- Layard, R., Clark, A. E., Cornaglia, F., Powdthavee, N., & Vernoit, J. (2014). What predicts a successful life? A life-course model of well-being. *Economic Journal*, 124, F720–F738. <http://dx.doi.org/10.1111/eoj.12170>
- LeCroy, C. W., & Krysik, J. (2008). Predictors of academic achievement and school attachment among hispanic adolescents. *Children & Schools*, 30, 197–209. <http://dx.doi.org/10.1093/cs/30.4.197>
- Li, Y., & Lerner, R. M. (2011). Trajectories of school engagement during adolescence: Implications for grades, depression, delinquency, and substance use. *Developmental Psychology*, 47, 233–247. <http://dx.doi.org/10.1037/a0021307>
- Libbey, H. P. (2004). Measuring student relationships to school: Attachment, bonding, connectedness, and engagement. *The Journal of School Health*, 74, 274–283. <http://dx.doi.org/10.1111/j.1746-1561.2004.tb08284.x>
- Malti, T., Ribeaud, D., & Eisner, M. P. (2011). The effectiveness of two universal preventive interventions in reducing children's externalizing behavior: A cluster randomized controlled trial. *Journal of Clinical Child and Adolescent Psychology*, 40, 677–692. <http://dx.doi.org/10.1080/15374416.2011.597084>
- Mashford-Scott, A., Church, A., & Tayler, C. (2012). Seeking children's perspectives on their wellbeing in early childhood settings. *International Journal of Early Childhood*, 44, 231–247. <http://dx.doi.org/10.1007/s13158-012-0069-7>
- Mavroveli, S., Petrides, K. V., Rieffe, C., & Bakker, F. (2007). Trait emotional intelligence, psychological well-being and peer-rated social competence in adolescence. *British Journal of Developmental Psychology*, 25, 263–275. <http://dx.doi.org/10.1348/026151006X118577>
- McIntosh, K., Mercer, S. H., Nese, R. N. T., Strickland-Cohen, M. K., & Hoselton, R. (2016). Predictors of sustained implementation of school-wide positive behavioral interventions and supports. *Journal of Positive Behavior Interventions*, 18, 209–218. <http://dx.doi.org/10.1177/1098300715599737>
- Mead, S., Hilton, D., & Curtis, L. (2001). Peer support: A theoretical perspective. *Psychiatric Rehabilitation Journal*, 25, 134–141. <http://dx.doi.org/10.1037/h0095032>
- Meyer, L. E., & Ostrosky, M. M. (2018). Identifying classroom friendships: Teachers' confidence and agreement with children. *Topics in Early Childhood Special Education*, 38, 94–104. <http://dx.doi.org/10.1177/0271121418763543>
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 2693–2698. <http://dx.doi.org/10.1073/pnas.1010076108>
- Monahan, K. C., Guyer, A. E., Silk, J., Fitzwater, T., & Steinberg, L. (2016). Integration of developmental neuroscience and contextual approaches to the study of adolescent psychopathology. In D. Cicchetti (Ed.), *Developmental psychopathology* (3rd ed., Vol. 2, pp. 720–765). Hoboken, NJ: Wiley. <http://dx.doi.org/10.1002/9781119125556.devpsy219>
- Morris, P., Mattera, S. K., Castell, N., Bangser, M., Bierman, K. L., & Raver, C. (2014). *Impact findings from the Head Start CARES demonstration* (OPRE Report 2014–44). Retrieved from <https://files.eric.ed.gov/fulltext/ED546649.pdf>
- Muthén, L. K., & Muthén, B. O. (1998–2017a). *Mplus user's guide* (8th ed.). Los Angeles, CA: Author.
- Muthén, L. K., & Muthén, B. O. (1998–2017b). *Warning: One or more individual-level variables have no variation within a cluster for the following clusters*. Retrieved from <https://www.statmodel.com/download/WARNING%20One%20or%20more%20individual-level%20variables%20have%20no%20within-cluster%20variance>
- Nagengast, B., Brisson, B. M., Hulleman, C. S., Gaspard, H., Häfner, I., & Trautwein, U. (2018). Learning more from educational intervention studies: Estimating complier average causal effects in a relevance intervention. *Journal of Experimental Education*, 86, 105–123. <http://dx.doi.org/10.1080/00220973.2017.1289359>
- Nangle, D. W., Erdley, C. A., Newman, J. E., Mason, C. A., & Carpenter, E. M. (2010). Popularity, friendship quantity, and friendship quality: Interactive influences on children's loneliness and depression. *Journal of Clinical Child and Adolescent Psychology*, 32, 546–555. [http://dx.doi.org/10.1207/S15374424JCCP3204\\_7](http://dx.doi.org/10.1207/S15374424JCCP3204_7)
- Novak, M., Mihić, J., Bašić, J., & Nix, R. L. (2017). PATHS in Croatia: A school-based randomised-controlled trial of a social and emotional learning curriculum. *International Journal of Psychology*, 52, 87–95. <http://dx.doi.org/10.1002/ijop.12262>
- Oberle, E. (2018). Early adolescents' emotional well-being in the classroom: The role of personal and contextual assets. *The Journal of School Health*, 88, 101–111. <http://dx.doi.org/10.1111/josh.12585>
- Oldfield, J., Humphrey, N., & Hebron, J. (2016). The role of parental and peer attachment relationships and school connectedness in predicting adolescent mental health outcomes. *Child and Adolescent Mental Health*, 21, 21–29. <http://dx.doi.org/10.1111/camh.12108>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716. <http://dx.doi.org/10.1126/science.aac4716>
- Panayiotou, M., Humphrey, N., & Wigelsworth, M. (2019). An empirical basis for linking social and emotional learning to academic performance. *Contemporary Educational Psychology*, 56, 193–204. <http://dx.doi.org/10.1016/j.cedpsych.2019.01.009>

- Parker, J. G., & Asher, S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology, 29*, 611–621. <http://dx.doi.org/10.1037/0012-1649.29.4.611>
- Payne, A. A., & Eckert, R. (2010). The relative importance of provider, program, school, and community predictors of the implementation quality of school-based prevention programs. *Prevention Science, 11*, 126–141. <http://dx.doi.org/10.1007/s11121-009-0157-6>
- Payne, A. A., Gottfredson, D. C., & Gottfredson, G. D. (2006). School predictors of the intensity of implementation of school-based prevention programs: Results from a national study. *Prevention Science, 7*, 225–237. <http://dx.doi.org/10.1007/s11121-006-0029-2>
- Pepler, D. J., & Craig, W. M. (2003). Assessing children's peer relationships. *Child Psychology & Psychiatry Review, 3*, 176–182. <http://dx.doi.org/10.1111/1475-3588.00238>
- Peugh, J. L., Strotman, D., McGrady, M., Rausch, J., & Kashikar-Zuck, S. (2017). Beyond intent to treat (ITT): A complier average causal effect (CACE) estimation primer. *Journal of School Psychology, 60*, 7–24. <http://dx.doi.org/10.1016/j.jsp.2015.12.006>
- Peugh, J. L., & Toland, M. D. (2017). Psychometric and quantitative methods for school psychology. *Journal of School Psychology, 60*, 5–6. <http://dx.doi.org/10.1016/j.jsp.2017.01.001>
- Rajmil, L., López, A. R., López-Aguilá, S., & Alonso, J. (2013). Parent-child agreement on health-related quality of life (HRQOL): A longitudinal study. *Health and Quality of Life Outcomes, 11*, 101. <http://dx.doi.org/10.1186/1477-7525-11-101>
- Ravens-Sieberer, U., Auquier, P., Erhart, M., Gosch, A., Rajmil, L., Bruil, J., . . . the European KIDSCREEN Group. (2007). The KIDSCREEN-27 quality of life measure for children and adolescents: Psychometric results from a cross-cultural survey in 13 European countries. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 16*, 1347–1356. <http://dx.doi.org/10.1007/s11136-007-9240-2>
- Ripski, M. B., & Gregory, A. (2009). Unfair, unsafe, and unwelcome: Do high school students' perceptions of unfairness, hostility, and victimization in school predict engagement and achievement? *Journal of School Violence, 8*, 355–375. <http://dx.doi.org/10.1080/15388220903132755>
- Robitail, S., Ravens-Sieberer, U., Simeoni, M. C., Rajmil, L., Bruil, J., Power, M., . . . the KIDSCREEN Group. (2007). Testing the structural and cross-cultural validity of the KIDSCREEN-27 quality of life questionnaire. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 16*, 1335–1345. <http://dx.doi.org/10.1007/s11136-007-9241-1>
- Robitail, S., Simeoni, M. C., Ravens-Sieberer, U., Bruil, J., Auquier, P., & Group, K. (2007). Children proxies' quality-of-life agreement depended on the country using the European KIDSCREEN-52 questionnaire. *Journal of Clinical Epidemiology, 60*, 469–478. <http://dx.doi.org/10.1016/j.jclinepi.2006.09.007>
- Roeser, R. W., van der Wolf, K., & Strobel, K. R. (2001). On the relation between social-emotional and school functioning during early adolescence: Preliminary findings from Dutch and American samples. *Journal of School Psychology, 39*, 111–139. [http://dx.doi.org/10.1016/S0022-4405\(01\)00060-7](http://dx.doi.org/10.1016/S0022-4405(01)00060-7)
- Ross, K. M., & Tolan, P. (2018). Social and Emotional Learning in Adolescence: Testing the CASEL Model in a normative Sample. *The Journal of Early Adolescence, 38*, 1170–1199. <http://dx.doi.org/10.1177/0272431617725198>
- Ross, S. M., Sheard, M. K., Cheung, A., Elliott, L., & Slavin, R. (2011). Promoting primary pupils' social-emotional learning and pro-social behaviour: Longitudinal evaluation of the Together 4 All Programme in Northern Ireland. *Effective Education, 3*, 61–81. <http://dx.doi.org/10.1080/19415532.2012.665773>
- Sagarin, B. J., West, S. G., Ratnikov, A., Homan, W. K., Ritchie, T. D., & Hansen, E. J. (2014). Treatment noncompliance in randomized experiments: Statistical approaches and design issues. *Psychological Methods, 19*, 317–333. <http://dx.doi.org/10.1037/met0000013>
- Schochet, P. Z., & Chiang, H. S. (2011). Estimation and identification of the complier average causal effect parameter in education RCTs. *Journal of Educational and Behavioral Statistics, 36*, 307–345. <http://dx.doi.org/10.3102/1076998610375837>
- Schonfeld, D. J., Adams, R. E., Fredstrom, B. K., Weissberg, R. P., Gilman, R., Voyce, C., . . . Speese-Linehan, D. (2015). Cluster-randomized trial demonstrating impact on academic achievement of elementary social-emotional learning. *School Psychology Quarterly, 30*, 406–420. <http://dx.doi.org/10.1037/spq0000099>
- Schultz, B. K., Evans, S. W., Langberg, J. M., & Schoemann, A. M. (2017). Outcomes for adolescents who comply with long-term psychosocial treatment for ADHD. *Journal of Consulting and Clinical Psychology, 85*, 250–261. <http://dx.doi.org/10.1037/ccp0000172>
- Sedgwick, P. (2015). Intention to treat analysis versus per protocol analysis of trial data. *British Medical Journal, 350*, h681. <http://dx.doi.org/10.1136/bmj.h681>
- Sklad, M., Diekstra, R., Ritter, M. D., Ben, J., & Gravesteyn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools, 49*, 892–909. <http://dx.doi.org/10.1002/pits.21641>
- Social Character Development Research Consortium. (2010). *Efficacy of school-wide programs to promote social and character development and reduce problem behavior in elementary school children: Report from the social and character development research program* (NCER 2011–2001). Retrieved from <https://ies.ed.gov/ncer/pubs/20112001/pdf/20112001.pdf>
- Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A., & Janssens, J. M. A. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review, 13*, 254–274. <http://dx.doi.org/10.1007/s10567-010-0071-2>
- Stormshak, E. A., Connell, A. M., Véronneau, M. H., Myers, M. W., Dishion, T. J., Kavanagh, K., & Caruthers, A. S. (2011). An ecological approach to promoting early adolescent mental health and social adaptation: Family-centered intervention in public middle schools. *Child Development, 82*, 209–225. <http://dx.doi.org/10.1111/j.1467-8624.2010.01551.x>
- Stuart, E. A., Perry, D. F., Le, H. N., & Jalongo, N. S. (2008). Estimating intervention effects of prevention programs: Accounting for noncompliance. *Prevention Science, 9*, 288–298. <http://dx.doi.org/10.1007/s11121-008-0104-y>
- Tanner-Smith, E. E., Durlak, J. A., & Marx, R. A. (2018). Empirically based mean effect size distributions for universal prevention programs targeting school-aged youth: A review of meta-analyses. *Prevention Science, 19*, 1091–1101. <http://dx.doi.org/10.1007/s11121-018-0942-1>
- Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development, 88*, 1156–1171. <http://dx.doi.org/10.1111/cdev.12864>
- Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 55–66). London, UK: National Foundation for Educational Research.
- Upton, P., Lawford, J., & Eiser, C. (2008). Parent-child agreement across child health-related quality of life instruments: A review of the literature. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 17*, 895–913. <http://dx.doi.org/10.1007/s11136-008-9350-5>
- Van Ryzin, M. J., Stormshak, E. A., & Dishion, T. J. (2012). Engaging parents in the family check-up in middle school: Longitudinal effects on

- family conflict and problem behavior through the high school transition. *Journal of Adolescent Health, 50*, 627–633. <http://dx.doi.org/10.1016/j.jadohealth.2011.10.255>
- Véronneau, M. H., Dishion, T. J., Connell, A. M., & Kavanagh, K. (2016). A randomized, controlled trial of the family check-up model in public secondary schools: Examining links between parent engagement and substance use progressions from early adolescence to adulthood. *Journal of Consulting and Clinical Psychology, 84*, 526–543. <http://dx.doi.org/10.1037/a0040248>
- Voegler-Lee, M. E., Kupersmidt, J. B., Field, S., & Willoughby, M. T. (2012). Student characteristics as predictors of teachers' implementation of a kindergarten readiness program. *Prevention Science, 13*, 472–482. <http://dx.doi.org/10.1007/s11121-012-0274-5>
- Wallander, J. L., & Koot, H. M. (2016). Quality of life in children: A critical examination of concepts, approaches, issues, and future directions. *Clinical Psychology Review, 45*, 131–143. <http://dx.doi.org/10.1016/j.cpr.2015.11.007>
- Wang, M. T., & Holcombe, R. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal, 47*, 633–662. <http://dx.doi.org/10.3102/0002831209361209>
- Weare, K., & Markham, W. (2005). What do we know about promoting mental health through schools? *Promotion & Education, 12*, 118–122. <http://dx.doi.org/10.1177/10253823050120030104>
- Weissberg, R. P., Durlak, J. A., Domitrovich, C. E., & Gullotta, T. P. (2015). Social and emotional learning: Past, present and future. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 3–19). New York, NY: Guilford Press.
- Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development, 62*, 1066–1078. <http://dx.doi.org/10.2307/1131152>
- Wigelsworth, M., Lendrum, A., Oldfield, J., Scott, A., Ten-Bokkel, I., Tate, K., & Emery, C. (2016). The influence of trial stage, developer involvement and international transferability on universal social and emotional learning programme outcomes: A meta-analysis. *Cambridge Journal of Education, 46*, 347–376. <http://dx.doi.org/10.1080/0305764X.2016.1195791>
- World Health Organization. (2003). *Creating an environment for emotional and social well-being: An important responsibility of a health promoting and child-friendly school*. Retrieved from <https://apps.who.int/iris/handle/10665/42819>
- Young Minds & National Children's Bureau. (2017). *Wise up: Prioritising wellbeing in schools*. Retrieved from <https://youngminds.org.uk/media/1428/wise-up-prioritising-wellbeing-in-schools.pdf>
- Zeanah, C. H., Berlin, L. J., & Boris, N. W. (2011). Practitioner review: Clinical applications of attachment theory and research for infants and young children. *Journal of Child Psychology and Psychiatry, 52*, 819–833. <http://dx.doi.org/10.1111/j.1469-7610.2011.02399.x>
- Zins, J. E., Bloodworth, M. R., Weissberg, R. P., & Walberg, H. J. (2004). The scientific base linking social and emotional learning to school success. In J. E. Zins, R. P. Weissberg, M. C. Wang, & H. J. Walberg (Eds.), *Building academic success on social and emotional learning* (pp. 3–22). New York, NY: Teachers College Press.

Received October 11, 2018

Revision received March 5, 2019

Accepted March 5, 2019 ■