Prevention Science (2018) 19:1066–1078
https://doi.org/10.1007/s11121-018-0948-8

How Program Users Enhance Fidelity: Implementing the WITS Programs in Rural Canadian Elementary Schools

Bonnie Leadbeater[1] · Kara Thompson[2] · Paweena Sukhawathanakul[1] · Gabriel J. Merrin[1]

Published online: 26 September 2018

Abstract

Research on the processes that enhance implementation fidelity is needed to increase understanding of ways to advance the uptake and sustainability of evidence-based programs (Berkel et al. in *Prevention Science, 12*, 23–33, 2011; Berkel et al. 2017). We propose and test a theoretical model of interrelations among implementation *fidelity* (i.e., adherence to program components), and rarely investigated activities of program users that may enhance fidelity; namely *integration* of program strategies into daily activities and *children's use of program strategies* (CUoPS). These were assessed across the initial 2 years of the implementation of the WITS peer victimization prevention programs in 16 Canadian rural schools. WITS stands for Walk away, Ignore, Talk it out, and Seek help. All schools were implementing the program. We examined the interrelations among these implementation indicators and their effects on child outcomes targeted by the WITS Programs (i.e., social responsibility, prosocial leadership, peer victimization, emotional symptoms, and aggression). Four implementation assessments were collected from children ($n = 1326$), their parents, and teachers in the fall and spring of two academic years. The within-time correlations among fidelity, integration, and CUoPS were significant at each assessment. Cross-lagged models showed fidelity and integration, and CUoPS and integration were reciprocally related during each academic year. CUoPS predicted higher subsequent levels of fidelity during and across academic years. Suggestions are given for training and coaching to enhance teachers' integration of program strategies into daily life and for creating opportunities for school staff to observe children using program strategies in order to enhance implementation fidelity.

Keywords Implementation · Fidelity · Implementation outcomes · WITS Programs

Evaluations of evidence-based social emotional learning (SEL) programs and bullying prevention programs in schools consistently conclude that quality of implementation strongly affects targeted child outcomes (see review by Taylor et al. 2017). Indeed, implementation outcomes, as distinct from treatment or intervention outcomes, have themselves become the focus of considerable theoretical and research attention. Implementation theory suggests that the activities involved in the large-scale implementation of evidence-based interventions (EBIs) (i.e., pre-adoption and adoption, implementation in real-world trials, and sustainability) are characterized by the increasing uptake and embedding of program strategies into real-world or everyday practices (Indig et al. 2017; Ogden and Fixsen 2014; Proctor et al. 2013). Reviewing implications of the expanding field of implementation science for schools, Forman et al. (2013, p. 78) define implementation broadly as "the process of putting a practice or program in place in the functioning of an organization, such as a school, and can be viewed as the set of activities designed to accomplish this." Although considerable research exists that identifies obstacles to implementation in schools (Berkel et al. 2011; Ogden and Fixsen 2014), we know less about how program users foster implementation fidelity. This study assesses users' efforts to embed program strategies into everyday practices and to observe program effects on student behaviors. We also examine whether these strategies impact implementation fidelity and early evidence of expected program (treatment) outcomes.

Several theoretical frameworks conceptualize the steps between becoming aware of an evidence-based program and its eventual incorporation into routine practices (e.g., see reviews

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11121-018-0948-8) contains supplementary material, which is available to authorized users.

✉ Bonnie Leadbeater
bleadbea@uvic.ca

[1] Department of Psychology, University of Victoria, 3800 Finnerty Rd, Victoria, British Columbia V8P 5C2, Canada

[2] Department of Psychology, St. Francis Xavier University, Antigonish B2G 2W5, Nova Scotia, Canada

by Greenberg et al. 2005; Ogden and Fixsen 2014; Proctor et al. 2009). Based on several of these frameworks, Proctor et al. (2011, p. 65) created a taxonomy of implementation outcomes that attempts to incorporate the multiple terms that are already beginning to spawn in the implementation science literature. Proctors et al.'s terms are *acceptability* (satisfaction), *adoption* (initial implementation), *appropriateness* (fit, relevance), *feasibility* (suitability for everyday use, practicability), *fidelity* (implemented as intended), *penetration* (reach, numbers served), and *sustainability* (integration, institutionalization, routinization). Theorists have also argued that these implementation outcomes can be conceptualized as a helix of interactive processes that need to begin early and be sustained across multiple years (Forman et al. 2013; Han and Weiss 2005; Leadbeater et al. 2015; Ogden and Fixsen 2014). Implementation research involving SEL programs has, with few exceptions, focused on factors affecting acceptability, adoption, and fidelity; however, less is known about the interactive processes sustaining program fidelity across the early years of implementation.

In this study, we propose a theoretical model of interrelations among implementation fidelity (i.e., adherence to program components), and users' *integration* of program strategies into daily activities, and *children's use of program strategies* (CUoPS). We examine the interrelations among these processes across the initial 2 years of the implementation of the evidence-based WITS peer victimization prevention programs. WITS stands for Walk away, Ignore, Talk it out, and Seek help. The WITS Programs aim to enhance children's social responsibility (e.g., caring for others) and prosocial leadership (e.g., takes initiative to help others) and to reduce peer victimization and related emotional problems and aggression. Support for this theory of change is presented in Leadbeater et al. (2016). The WITS Programs have also been evaluated in two quasi-experimental studies (Giesbrecht et al. 2011; Hoglund et al. 2012; Leadbeater and Sukhawathanakul 2011). Program schools show small, but significantly more rapid declines, in peer physical and relational victimization compared to control schools. To reduce inequities to evidence-based programs in rural school districts, training modules and program resources for school staff, parents, community leaders, and children can be accessed online at www.witsprogams.ca. In contrast to manual-based programs in which facilitators follow prescribed lesson plans, the WITS Programs take a whole school approach designed to engage multiple users in creating a new whole school approach to reducing peer victimization and bullying. As such, program fidelity requires the daily use of the program strategies (see Hansen 2014). We also test the effects of theoretical model on child treatment outcomes targeted by the WITS Programs (i.e., social responsibility, prosocial leadership, peer victimization, emotional symptoms, and aggression).

Situating Sustainability Planning in Ongoing Implementation Outcomes

Implementation science has illuminated multiple systemic conditions that support or impede initial implementation steps in schools; such as, organizational policies, readiness for change, supportive attitudes, burnout, necessary initial and ongoing contextual supports, funding, and leadership (Greenberg et al. 2005; Forman et al. 2013). These conditions impact decisions about the overall acceptability, adoption, and appropriateness of a program. Somewhat surprisingly, the users' (e.g., teachers, parents, students, and administrators) day-to-day implementation activities (integration, institutionalization, and routinization) are somewhat hidden in the "black box" of implementation (Ogden and Fixsen 2014, p. 5). These naturalistic variations in implementation activities are unlikely to be captured by standard indicators of program feasibility or fidelity because they reflect how program strategies are gradually put to use in the course of daily practice, rather than users' adherence to manualized activities (Forman et al. 2013).

Implementation success may not only be the result of adherence to manualized activities, but also require and interact with efforts on the part of users to personalize and incorporate the program principles and strategies into their daily practices and to notice the impact of their effort on changing children's behaviors. For example, Berkel et al. (2011) describe a model of how practitioners (e.g., school staff) and clients (children) may contribute to implementation outcomes: They suggest that competent, clear, and enthusiastic presentation of the program (e.g., interactive teaching) will affect children's engagement and response and subsequently enhance treatment outcomes. Also, outlining a model examining the sustainability of teacher implementation of school-based mental health programs in schools, Han and Weiss (2005, p. 674) suggest that sustainability processes unfold in all phases of implementation (i.e., pre-implementation, supported implementation, and sustaining practices when external funding has been withdrawn). In particular, the authors argue that observations that allow teachers to attribute improved student functioning to the program activities, which they have enacted, operate at the level of teachers and children to create potentially self-sustaining feedback loops that, over time, can result in high implementation fidelity and quality and, eventually, positive changes in child treatment outcomes. Han and Weiss (2005) argue that "teachers' *generalized use of effective program techniques,* which reflects a broader application of program techniques and strategies reflecting core program principles to various classroom situations" (p. 675) is a central to program fidelity and sustainability. Yet, the integration of program strategies into daily activities is typically left up to the users as they adopt and adapt manualized strategies to fit their needs

(Owens et al. 2014) and a better understanding of how to promote the incorporation of program strategies to everyday practice is important to sustaining motivating to use the program with fidelity.

Little research has identified or directly examined how variability in these everyday activities affects implementation fidelity or treatment outcomes. In a two-and-a-half-year longitudinal qualitative study, Leadbeater et al. (2015) interviewed school staff involved in the implementation of the WITS Programs in eight elementary schools. Participants described processes that fostered self-sustaining cycles of engagement over time (e.g., building on staff buy-in and embedding the program activities, using the WITS common language in everyday practice, and institutionalization of WITS in school codes of conduct). Staff also noted a need for annual self-reflection and renewal of commitments to using the programs. The authors conclude that sustainability planning in schools involves multi-level, ongoing processes that need to be anticipated and supported by school leadership, program champions, and developers to help schools to realize their investment in starting to use evidence-based programs.

Also, using data from year one of a randomized controlled trial of Second Step® involving 61 schools, Low et al. (2016) identified two indicators that demonstrated teachers' skill acquisition beyond adherence to manualized program activities: (1) teachers' reinforcement and integration of the program strategies into daily activities, which we will call "integration" and (2) observations of children's use of program strategies in everyday behaviors, which we call "children's use of program strategies" (CUoPS). Low et al. (2016) found that these were evident in a high-implementation class of schools, but were rare in schools where teachers also showed poor fidelity to the manualized program. The authors concluded that these "less scripted dimensions (i.e., aspects that are harder to manualize) are necessary to obtain maximum benefits from program implementation" (Low et al. 2016, p.989). However, the interactions among and changes in these processes were not the focus of the Low et al. (2016) study, and these indicators were not assessed over time.

The current study extends previous theory and research on process that sustain implementation fidelity by focusing on the efforts of program users to use program strategies in daily practices. In Fig. 1, we specify our model suggesting that integration into daily practice and CUoPS interact with fidelity in program implementation efforts (Berkel et al. 2011; Han and Weiss 2005; Leadbeater et al. 2015; Wanless and Domitrovich 2015). In addition, high initial levels of implementation fidelity may be needed to increase the likelihood of observing student behaviors that can be attributed to the program, and behavioral changes on the part of children likely precede successful treatment outcomes (Andreou et al. 2015).

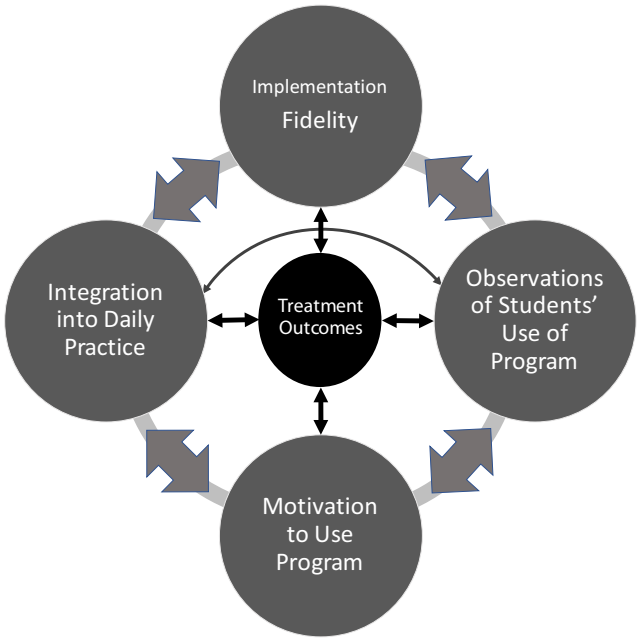

Fig. 1 Theoretical model of relations of implementation fidelity, teachers' efforts to integrate program activities in daily practices, and observations of children's use of program with targeted treatment outcomes

Assessing Implementation Outcomes in Real-World Studies

Ogden and Fixsen (2014) note the many methodological challenges encountered in implementation research and the need to operationalize and develop measures of new components. In EBIs that engage multiple implementers, the assessment of implementation processes can be further complicated by the need to examine several implementation processes across multiple reporters (e.g., administrators, teachers, parents, and children). Moreover, repeated assessments are needed to examine how implementation processes are sustained and interact over time. Ratings of adherence to program lessons are often provided by the individuals (often classroom teachers) who are delivering the lessons or, more ideally, by independent observers of each lesson who are trained by research teams (e.g., Smith et al. 2018). However, Procter et al. (2013) argue that "measures used [to assess implementation outcomes] in efficacy research may prove too cumbersome for real-world studies" (p. 72). Teacher ratings can be biased by self-interest and incomplete data. However, observer ratings can be impractical, costly, and resisted by school staff who are too busy or feel that they are being evaluated (Dusenbury et al. 2004). Establishing and sustaining inter-observer reliability across wide distances, and changing school or research staff are additional challenges (Hirschstein et al. 2007; Smith et al. 2018). In this study, we address potential bias in reporting of implementation activities by examining indices of implementation process that reflect responses of children, parents, and teachers.

Summary

We examine processes that may impact implementation fidelity in the course of an effectiveness evaluation of the WITS Programs in which all school were implementing the programs. The programs were being implemented in 16 elementary schools in rural or suburban school districts from Alberta, Ontario, and New Brunswick. We assessed indices of implementation fidelity as well as teacher integration of program strategies into their daily practice, and children's use of WITS strategies (CUoPS) four times over two academic years. We examine the stability of and interactions among fidelity, integration, and CUoPS and also examine their impact on intervention outcomes targeted by the WITS Programs.

Method

Participants

Children ($n = 1967$; 1027 female), their parents, and teachers were recruited in 2011 from rural school districts in three provinces (Ontario, Alberta, and New Brunswick) to evaluate the large-scale implementation of the WITS Programs (see Leadbeater et al. 2016 for details). Schools in each district were randomly assigned to program implementation ($n = 16$, including 1329 children at baseline) or control groups ($n = 11$, including 638 children at baseline). Program schools were oversampled to enable examination of expected differences in implementation quality. Baseline data were collected from grades 1 to 3 in order to have some of the children progress to the WITS LEADerS curriculum (which included activities that are developmentally appropriate for grades 4 to 6). All children would have received the WITS Program for 2 years after baseline.

Participating children and their parents and teachers each completed surveys five times: at baseline (T1 spring prior to implementation) and in the fall and spring over the two subsequent academic years (T2 to T5). Implementation data were collected in program schools only at T2 to T5. Children participating in each wave and the numbers with parent and teacher data are shown in Fig. 2. Fifty children and their parent and teachers were added at T2 due to delays in obtaining parent consent at the baseline assessment. Overall, children's participation remained high (96% at T5). However, participation by parents declined over time from 80 to 54%. The teachers (range $n = 70$–75 at each assessment) provided data for the majority of participating children (76 to 82%).

Parent reports indicated that the majority (93%) of participants in the program schools were European Caucasian, 2% Aboriginal, 1% Asian, 2% were of mixed ethnicity, and less than 1% were Hispanic, Indo-Canadian, or African Canadian. Most (73%) children lived in a two-parent household and most (70%) had attended their current school only. Socioeconomic status (SES) was diverse: 25% of mothers and 32% of fathers had high school education, whereas 45% of mothers and 46% of fathers completed some 2-year college or technical training and 30% of mothers and 22% of fathers had some university education beyond high school. Selective attrition of children was assessed by testing for differences in T1 demographic variables (sex, age, maternal and paternal education) between program children who remained in the longitudinal study ($n = 949$) at T5 and those who had no data at T5 ($n = 186$). No differences were significant.

Procedure

Implementation Training

Trained research assistants were overseen by PhD-level researchers at each program school. In the fall of academic year 1 (T2), the research assistants guided school staff and administrators through standardized online training modules in groups of 20 to 40 on a professional development day. Teachers who completed the training received a certificate of accreditation that was submitted to the research staff by the principal. Nonprofessional school staff and classroom assistants attended this training and contributed to the discussion. To accommodate staff turnover, at the beginning of academic year two (T4), teachers new to the school completed the online training modules individually with the guidance of the research assistants. Community leaders were recruited by school district superintendents and principals from local detachments of the Royal Canadian Mounted Police (RCMP), and each RCMP member completed standardized online training modules and obtained a certificate of accreditation. "WITS gifts" (e.g., rulers, pencils, erasers, fridge magnets) given by the community leaders during classroom visits were also sent home to remind parents to use the program language and strategies. Pamphlets explaining the WITS Programs language and strategies were sent home to parents by the research team with the consent materials and questionnaires.

Program schools received all supplies needed to implement the program including approximately 40 popular and publicly available children's books, classroom posters, and take-home gifts. Research staff observed book-based WITS lessons delivered by teachers once per year to provide encouragement and coaching. Also, 3 or 4 months after each wave of data collection, the school principals received summaries of their schools' levels of implementation of core components and levels of peer victimization reported by the children and their parents.

Data Collection

At each assessment, home room teachers sent home research packages including informed consent forms and questionnaires to parents of all children with instructions to return the completed or declined package in a sealed envelope. A mood-changing

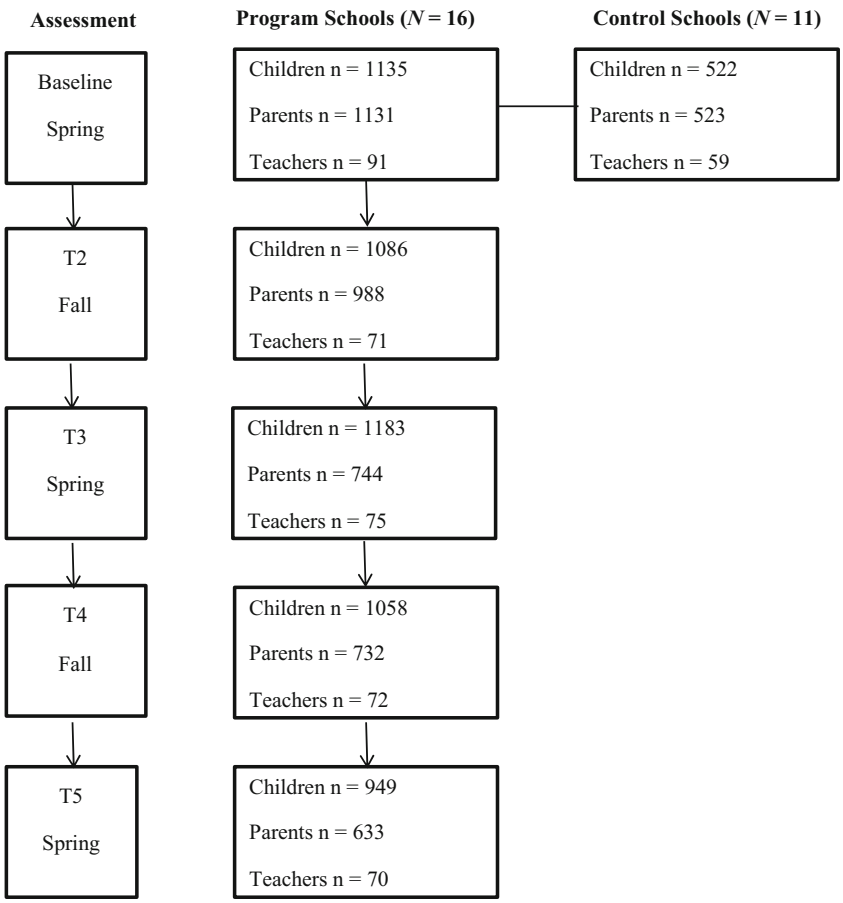

Fig. 2 Participant flow and data sources

pencil was given to students who returned the package whether or not consent was given. Schools recruited at least 65% of the eligible students (average of 75%). Teachers gave written consent for their own participation at each assessment and completed questionnaires (social responsibility in the classroom, leadership, aggression, emotional problems) for each child who had parent consent. Teachers received a gift certificate at each assessment. A parent or guardian (typically mothers) also rated children's prosocial leadership, emotional problems, and aggressive behaviors. Students completed questionnaires about their experiences with victimization in groups in regularly scheduled classroom time. Items were read aloud by trained research assistants and a second assistant circulated to ensure privacy and to respond to questions. Teachers, parents, and children also completed implementation items tapping the core program activities at T2 to T4.

Measures

Core Activities

The WITS Programs' core activities prescribed in the program manual are as follows: (1) Each school year is launched by a "Swearing in Ceremony" led by members of the RCMP. All students pledge to "Use their WITS to deal with peer conflict and to help others use their WITS." (2) Each month, teachers select books from an online list of more than 50 "WITS PICKs" and choose activities from lesson plans that support social emotional learnings (inclusion, kindness, courage, resolving conflicts) and are also integrated with academic objectives for learning English, Math, Social studies, Music, and Drama. (3) School staff and parents are encouraged to integrate "using your WITS" into daily practice using "teachable moments" to help children solve conflicts with their peers as they happen. (4) WITS Community leaders re-visit the school regularly to ask children about their successes in "using their WITS" and to deliver take home gifts (pencils, rulers, etc.) to support using WITS at home. (5) Children in grades 4 to 6 follow the WITS LEADerS lesson plans and learn to Look and Listen, Explore Points of View, Act, ask Did it Work, and Seek Help. This empowers them to help younger children in playgrounds, at schools, and in their communities. (6) Multilingual online resources and school and program outreach encourage parents to use the WITS language at home.

Fidelity, Integration of the Program Strategies into Daily Activities, and Children's Use of Program Strategies

To create assessment indices of fidelity, integration, and CUoPS, items were drawn from the questionnaires examining delivery

core program activities that were administered at each assessment to teachers, parents, and children (see Leadbeater et al. 2016; 1 = yes or 0 = no). These items were created by researchers who were observers of the program being used in all schools. Evidence-based programs need to create indices of program-specific activities and these are valid for the WITS Programs. As shown in Table 1, on average, endorsement of each of the items was adequate to high by T5 (51 to 87%), apart from one item assessing children's reports for "My family talks about WITS at home" (31–34%). Also as seen in Table 1, *fidelity* items indicate the degree to which facilitators complete program delivery as outlined in the online manual. *Integration* items tap activities, mainly by teachers that reinforce program strategies in everyday practices. *Children's use of programs strategies* (CUoPS) items indicate children's use of the WITS language at school, home, and neighborhoods. Items were assigned to each index by consensus ratings of the research team members and scores were summed within each index. Correlations between the indices were significant but small to moderate (see Table S1), indicating discriminant validity. As reported below, auto-regressions examined in path models showed significant stability across assessments (see Fig. 3a) for integration (.17–.19), fidelity (.17–.21), and CUoPS (.43–.45).

Program Outcomes

The *social responsibility* measure (Leadbeater and Sukhawathanakul 2011) reflects the curriculum objectives of the British Columbia Ministry of Education Performance Standards: Social Responsibility Framework for children in grades 1 to 6. Teachers rated each child on six items (e.g., "looks for chances to include others," and "helps to solve peer conflicts"). Teachers rated how well (compared to other children in the class) each child met their expectations for socially responsible behaviors measured on a four-point Likert scale (0 = *not yet within expectations*, to 3 = *exceeds expectations*). Items were summed. Cronbach's alphas ranged from .94 to .96.

Prosocial leadership was assessed using the leadership and prosocial competence items from the Behaviour Assessment System for Children (BASC; Reynolds and Kamphaus 2004). Parents and teachers rated eight items on a four-point scale (0 = "*hardly ever*," 1 = "*sometimes*," 2 = "*often*," 3 = "*almost always*") indicating how often children displayed leadership (e.g., "is good at getting people to work together") and prosocial competence (e.g., "offers to help other children"). Parent and teacher ratings were correlated at each time point (range = .29–.70). Cronbach's alphas ranged from .90 to .92 for the teacher reports and from .80 to .83 for the parent reports. To combine parent and teacher ratings, latent variables were created using item-level data.

Peer victimization was assessed with a version of the Social Experience Questionnaire (SEQ; Crick and Grotpeter 1996) adapted for younger students (Desjardins et al. 2013). Five items assessed child reports of *relational victimization* (e.g., "how often does another kid tell lies about you to make others not like you anymore?") and five items assessed *physical victimization* (e.g., "how often do you get pushed or shoved by another kid at school?"). Item responses were indicated in words and by small (never), medium (sometimes), and large (all most all the time) boxes on a three-point Likert scale (0 =

Table 1 Item percentages of implementation variables

Dimensions of implementation	T2	T3	T4	T5
Fidelity variable				
1. Read 3 or more books in class (teacher report)	39.3%	51%	40.2%	52.6%
2. Tug of help or swearing in (child report)	83.9%	87.7%	93%	87.9%
3. Classroom visitor (child report)	54.4%	78.8%	74%	77%
4. Parent received WITS pamphlet or newsletter (parent report)	81.2%	80.8%	80.5%	80.0%
Integration variable				
1. Recognized a student for WITS 3 or more times (teacher report)	66.9%	69%	67.7%	74.7%
2. Displayed student WITS projects 1 or more times (teacher report)	31.4%	49.3%	46.2%	64.7%
3. My teacher talks about WITS in the classroom (child report)	80.2%	84.4%	82.1%	78.2%
4. I brought home a WITS gift (child report)	65.9%	78.5%	66.6%	75%
5. I saw WITS posters in the school (parent report)	23.3%	43.9%	50.1%	56.7%
Children's use of program strategies variable				
1. I use my WITS to deal with bullying (child report)	82.3%	82.1%	82.5%	79.1%
2. My family talks about WITS at home (child report)	35.9%	37.8%	36.2%	31.2%
3. I use WITS with my friends in my neighborhood (child report)	59.8%	57.7%	58.1%	51.4%
4. I have helped other kids use their WITS (child report)	71.1%	69.9%	75.5%	69.6%
5. My child uses WITS language (parent report)	40.9%	62.8%	66.7%	70.0%

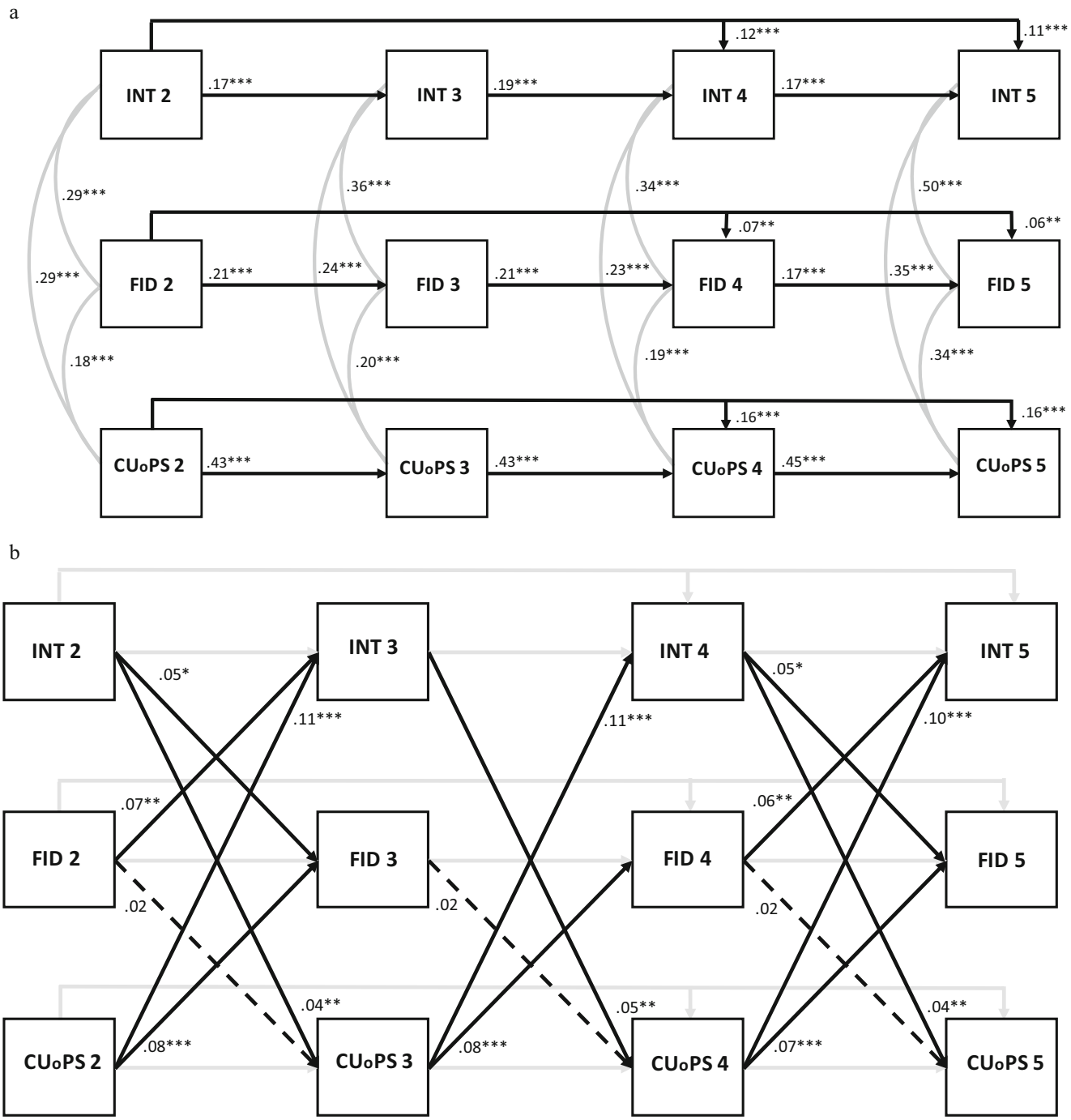

Fig. 3 **a** Presents auto-regressive and within-time correlations among dimensions of implementation over time and **b** presents relations among dimensions of implementation over time. *Note.* Gender, SES, and age at baseline are regressed on each dimension of implementation at each time point (paths not shown). INT integration, FID fidelity, CUoPS children's use of program strategies. Auto-regressive and cross-lagged paths are constrained across time. All estimates are standardized. Significant at $*p < .05$, $**p < .01$, $***p < .001$. Dashed lines indicate tested paths that are not significant. Within-time correlations are included but not shown in Fig. 2b

never, 1 = *sometimes*, and 2 = *almost all the time*). Responses were summed. Cronbach's alphas ranged from .86 to .89.

Aggression and emotional problems were assessed with items from the Early School Behavior Rating Scale (Caldwell and Pianta 1991). Parents and teachers rated on a four-point scale (0 = "*hardly ever*," 1 = "*sometimes*,"| 2 = "*often*," 3 = "*almost always*") how often children displayed *aggressive behaviors* (two items; "kicks, bites, or hits other children," "fights with other children"), and *emotional problems* (four items; "worries," "cries easily," "has headaches or stomach aches," "appears unhappy or depressed"). Internal consistency coefficients using polychoric correlations for ordinal scales

(Gadermann et al. 2012) ranged from .81 to .90 for teacher reports and from .77 to .83 for parent reports for aggression; and from .84 to .86 for teacher reports and from .69 to .77 for parent reports for emotional problems. To combine parent and teacher ratings, latent variables were created using item-level data for aggression and emotional problems.

Analyses Plans

To assess associations among the indices over time and their impact on program outcomes, we used auto-regressive cross-lag (ARCL) models (Little 2013) using MPlus 7.4 (Muthén and Muthén 1999–2012). Sex, SES, and age at baseline were regressed on fidelity, integration, CUoPS, and the outcome variable at each time point. Students were nested within classroom teacher at T2 and T3 and transitioned to a second teacher for T4 and T5. To account for both the change in teachers over time and the associated dependency related to data supplied by the teachers for more than one child, we estimated residual variance for each implementation and outcome variable for each specific teacher. To fit the cross-classified models, we used the Bayesian estimator available in Mplus 7.4. At T1, teacher-level variance made up 9% of the total variance for peer victimization, 3% for internalizing problems, 4% for aggression, and 1% for prosocial leadership. School-level variance made up 2% of the total variance for peer victimization, 1% for internalizing problems, 2% for externalizing problems, and 4% for prosocial leadership.

Relations Among Implementation Indices

We estimated auto-regressive paths and within-time correlations for the implementation indices (fidelity, integration, and CUoPS). Because initial levels of the indices were also expected to be associated with subsequent assessments, we added additional auto-regressive paths from fidelity, integration, and CUoPS at T2 to their respective assessments at T4 and T5. To examine the interactions among the implementation indices, we added cross-lagged pathways between fidelity and integration (only with each academic year due to changes in teachers) and with CUoPS at each assessment.

Effects of Implementation Indices on Child Treatments Outcomes

Using the cross-lagged model, we assessed how paths in the best fitting model of the associations among the fidelity, integration, and CUoPS were related to child program outcomes (social responsibility, prosocial leadership, victimization, emotional problems, and aggression). Due to power limitations, we examined each outcome variable separately. We specified lagged paths from fidelity, integration, and CUoPS to subsequent assessments of the outcome variable. We also examined mediation effects suggested by temporally sequenced significant paths (Hayes and Scharkow 2013; Muthén and Muthén 1999–2012). Attrition from T2 to T5 was 13% for children, 36% for parents, and 1% for teachers. For participants who contributed information at any one wave, missing data among study variables ranged from 1 to 8%. To reduce bias due to missing data, we used a multiple imputation strategy ($k = 100$) using the EM algorithm available in SAS 9.4. Under the missing at random (MAR) assumption, the imputation model using the expectation maximization algorithm provides unbiased estimates for missing data (McLachlan et al. 2004).

Results

Descriptive Statistics

Means and standard deviations for fidelity, integration, CUoPS, and child outcomes at each time point are presented in Table 2. Bivariate correlations among fidelity, integration, and CUoPS are presented in supplementary (S) Table S1, and correlations among the child outcomes are presented in Table S2.

Auto-Regressive Paths and Within-Time Correlations

Standardized estimates for auto-regression pathways are shown in Fig. 3a. The implementation indices were significantly and positively associated with each other within-time. Results show low to moderate stability for each of the indices between (and across) assessments. For example, a one standard deviation (SD) increase in CUoPS at T2 was associated with a .43 SD increase in CUoPS at T3. Significant covariates showed that CUoPS was higher for females than males, and fidelity, integration, and CUoPS were higher in younger than older students.

Cross-Lag Paths Model Examining Relations among Fidelity, Integration, and CUoPS

Cross-lagged pathways connecting the indices were added to the auto-regressive model. Standardized estimates for paths are shown in Fig. 3b. In support of our theory that integration and CUoPS affect and are affected by implementation fidelity, significant cross-lagged paths show that fidelity and integration were reciprocally related during each academic year (i.e., T2 toT3 and T4 toT5): During each of the academic years, fidelity was associated with increases in integration during ($\beta s = .06$ to .07), and integration was also associated with increases in fidelity ($\beta s = .05$). CUoPS also predicted increases in fidelity at each subsequent assessment ($\beta s = .07$ to .08), but was not predicted by fidelity. Both during and across (T3 toT4) the academic years, integration and CUoPS were also reciprocally related: Integration predicted CUoPS ($\beta s = .04$ to .05) and CUoPS predicted integration ($\beta s = .10$ to .11).

Table 2 Means and standard deviations of implementation and outcome variables over time

	Time 2		Time 3		Time 4		Time 5		*F*-test
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Integration	1.45^{a}	1.01	2.01^{bc}	1.05	1.90^{b}	0.98	2.11^{bc}	0.98	$F = 109.9$ $p < .001$
Fidelity	1.89^{a}	1.11	2.29^{b}	0.99	2.33^{b}	0.93	2.11^{c}	0.98	$F = 54.21$ $p < .001$
Children's use of program strategies	1.92^{a}	1.15	2.06^{a}	1.19	2.05^{a}	1.10	1.74^{b}	1.15	$F = 21.75$ $p < .001$
Social responsibility	9.50^{a}	4.13	9.75^{a}	4.19	9.94^{ab}	4.23	10.30^{b}	4.35	$F = 8.45$ $p < .001$
Prosocial leadership	12.32^{a}	3.62	12.31^{a}	3.78	12.16^{a}	3.78	12.28^{a}	4.03	$F = 0.51$ $p = .675$
Victimization	5.26^{a}	4.36	5.31^{a}	4.46	4.89^{ab}	4.18	4.58^{b}	3.90	$F = 8.78$ $p < .001$
Emotional problems	2.12^{a}	1.61	2.10^{a}	1.69	2.02^{a}	1.66	2.12^{a}	1.72	$F = 1.13$ $p = .337$

Note. Means with the same superscript are not significantly different. Means with different superscripts indicate significant mean differences. With Bonferroni correct for post hoc pairwise comparisons, the significance levels are set at $p < .01$

Associations Among Implementation Indices and Child Outcomes

Standardized estimates for the associations between each of the implementation indices (fidelity, integration, and CUoPS) and child outcomes (social responsibility, prosocial leadership, victimization, emotional problems, and aggression) are shown in Fig. 4. To enhance readability, paths for associations among fidelity, integration, and CUoPS are not shown. Within-time associations between the implementation indices and each outcome variable are also included in the models but are not shown (see Table S3).

Social Responsibility

Within-time associations were significant between higher integration and higher CUoPS and more social responsibility at T2, T3, and T4 (see Table S3). Within-time correlations with fidelity were not significant at any assessment. CUoPS predicted increases in social responsibility at each subsequent assessment ($\beta = .04$, see Fig. 4a). Mediation analyses also revealed that the longitudinal paths from integration to CUoPS to social responsibility were significant ($b = .01$, $SD = .006$, 95% confidence interval .001 to .023, see Table S4).

Emotional Problems

Within-time associations were significant between fidelity and CUoPS at T3 and T5 (see Table S3). Within-time correlations with integration were not significant at any assessment. As shown in Fig. 4b, CUoPS was associated with decreases in emotional problems at each subsequent assessment ($\beta = -.06$). Mediation analyses also revealed that the longitudinal path from integration to CUoPS to emotional problems was significant ($b = -.001$, $SD = .001$, 95% confidence interval $-.003$ to $-.000$, see Table S4).

Prosocial Leadership

Prosocial leadership was associated within-time with integration at T2, T4, and T5, with fidelity at T2, and with CUoPS at T2 and T5 (see Table S3). Cross-lagged effects were not significant (see Fig. S1a).

Victimization

The within-time associations between the implementation indices and victimization were not significant at any assessment (see Table S3). Cross-lagged effects were also not significant (see Fig. S1b).

Aggression

Within-time and cross-lagged associations between each of the implementation indices and aggression were not significant, except for one association between fidelity and aggression at T3. Cross-lagged effects were not significant (see Fig. S1c).

Discussion

There is considerable evidence that the success of evidence-based SEL interventions in improving children's behaviors is dependent on the quality of implementation (Taylor et al. 2017). Whole school interventions emphasize approaches that (a) integrate program strategies into children's daily activities, (b) sustain interventions across multiple years, (c) address broad risk and protective factors, and (d) engage individuals from

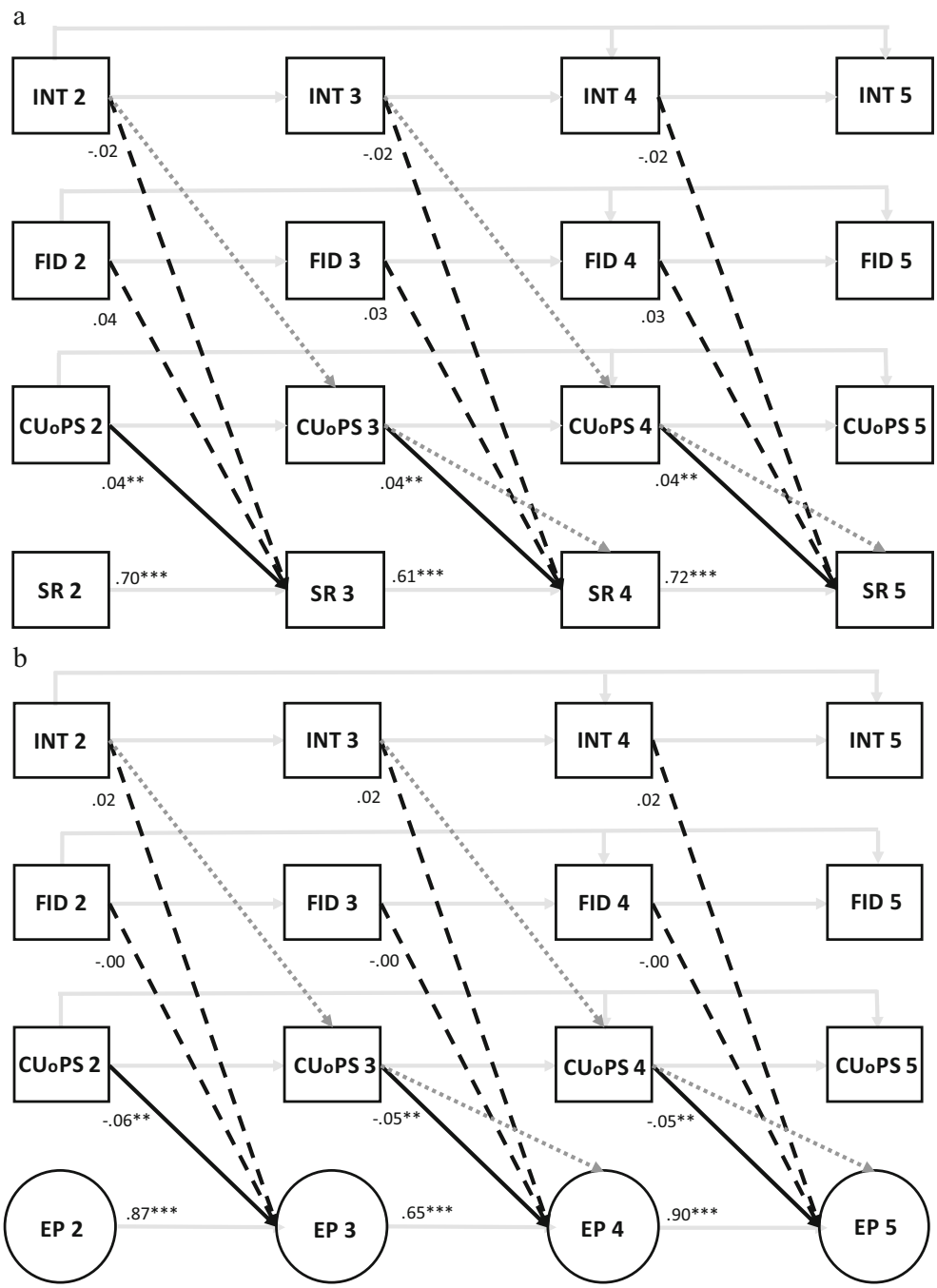

Fig. 4 The effects of implementation on social responsibility and emotion problems over time. *Note.* Gender, SES, and age at baseline are regressed on each dimension of implementation and outcome variable at each time point (paths not shown). All cross-lagged effects among implementation variables (see Fig. 2) and within-time correlations are included but note shown. INT integration, FID fidelity, CUoPS children's use of program strategies, SR social responsibility, EP emotion problems. Auto-regressive and cross-lagged paths are constrained across time. All estimates are standardized. Significant at $*p < .05$, $**p < .01$, $***p < .001$. Dashed lines indicate tested paths that are not significant. Solid gray lagged paths indicate significant mediation pathways

multiple key ecologies that surround children and youth (Greenberg et al. 2005). Hence, implementers of SEL programs need a better understanding of the processes and activities that impact implementation fidelity (adherence, dosage, quality) and also sustain school's investments in these programs across several academic years (Forman et al. 2013; Greenberg et al. 2005). Activities that can interact with, enhance, and sustain fidelity have been suggested by reviewers (Berkel et al. 2011; Han and Weiss 2005; Ogden and Fixsen 2014); however, these have not been the subject of much empirical research (but see Berkel et al. 2017). Odgen and Fixsen argue that the naturally occurring, user-driven activities that enhance fidelity are often hidden in the "black box" of implementation. These may also be left up to the school staff as they adopt and adapt manualized strategies to fit their own needs and expertise (Owens et al. 2014). Expanding on previous research and theory (Berkel et al. 2011; 20147; Han and Weiss 2005; Leadbeater et al. 2015), we suggest that implementation fidelity may motivate and be motivated by efforts to integrate program strategies into day-to-day practices and observations of children's use of program strategies (see Fig. 1). As Han and Weiss (2005) argue, day-to-day implementation activities can create self-reinforcing feedback loops that increase or

sustain fidelity and enhance treatment outcomes over time. Specifically, we found that implementation fidelity, teachers' efforts to integrate program strategies in daily practices, and observations of children's use of program language were related to each other and also to subsequent treatment outcomes in a 2-year study conducted in elementary schools who were all implementing the WITS Programs. All models accounted for nesting of children by teacher reports; including, both the change in teachers from one academic year to the next and the associated dependency related to data supplied by the teachers for more than one child. Replication of these WITS-specific findings in the context of other EBIs and with larger samples is needed; however, the findings provide insight into how fidelity, integration, and CUoPS may operate together to affect each other and treatment outcomes.

The Relations Among Fidelity, Integration, and Children's Use of Program Strategies

The within-time correlations among fidelity, integration, and CUoPS were significant at each assessment. Cross-lagged models showed fidelity and integration were reciprocally related in each academic year (i.e., children who experienced more fidelity also experienced more teacher integration of program strategies). CUoPS and integration were also reciprocally related, and CUoPS predicted higher subsequent levels of fidelity. Han and Weiss (2005) posited that incorporating program strategies into daily practice and witnessing children using program strategies (i.e., the WITS language) may motivate fidelity (adherence to the manualized program). It is possible that the reinforcement of WITS strategies in teachers' everyday practice increases the visibility of these strategies to children (beyond manualized lessons and activities) and CUoPS sets in motion greater interest both in adherence to the program and in integrating program strategies with everyday practices. Teachers' enthusiastic presentation of the program may enhance children's engagement and response and, subsequently, treatment outcomes.

More research is needed to better understand how to support integration and CUoPS in early years of program implementation. Our implementation training focused mainly on carrying out the prescribed program activities, and incorporation of program activities into everyday behaviors was, as is typical, left up to teachers, program champions, and administrators. The sophisticated skills needed for integrating program strategies into everyday practice may be less scripted and harder to manualize (Low et al. 2016). These taken-for-granted behaviors may rely on teacher expertise and enthusiasm and may be insufficiently emphasized in program development, initial training, and subsequent coaching, focused on adherence to manualized core program activities. According to Owens et al. (2014, p.111), a "train-and-hope" implementation model that is focused on session-oriented activities may not produce a change in implementers' or children's behaviors that are needed to improve treatment outcomes. Directed efforts to help teachers and schools' administrators explore the "fit" of EBIs with their own teaching styles, class management activities, activities, learning objectives, programs, and goals may be needed.

As theorized, fidelity was enhanced (beyond stability and within-time associations) by both integration and CUoPS. Our measure of integration mainly taps teachers' reports of reinforcing and recognizing the use of program strategies in their classrooms. Previous research notes the importance of teacher buy-in and expertise for readiness to adopt an evidence-based program (Han and Weiss 2005; Wanless and Domitrovich 2015) and adherence to program delivery (Low et al. 2016). It is possible that teachers' everyday use of program strategies reflect their buy-in, enthusiasm, and openness to change. Qualitative research also demonstrates the importance of school-staffs' attributions of changes in children's behaviors to program effects for fidelity (Leadbeater et al. 2015). Training and ongoing coaching could help program users explore the "fit" of program strategies with their own day-to-day teaching practices and to create opportunities (e.g., weekly class meetings or when reading WITS books) to ask children how they are using program strategies. Being able to observe indicators of students' use of strategies that can be attributed to the program may motivate increases in fidelity and integration. Helping program users to identify changes in children's activities that *they* expect as a result of program implementation (e.g., handling some playground conflicts independently) may also enhance teachers' perceptions that the program has positive effects (Han and Weiss 2005) and sustain fidelity and integration within and across grades. In the WITS Programs, community leaders, principals, and teachers are encouraged to ask children about their successes in "Using their WITS"—eliciting observations of CUoPS that may motivate their use of the programs.

Effects of the Implementation Indices on Targeted Outcomes

We also used ARCL models to examine the effects of fidelity, integration, and CUoPS on each of the child outcomes (i.e., social responsibility, prosocial leadership, victimization, and emotional problems) while considering the nested nature of children within multiple teachers. All models examining outcomes also include the auto-regressive associations, within-time correlations, and cross-lagged relations among each of the implementation indices. Consistent with previous research with this sample in which all children were in program schools (Leadbeater et al. 2016), CUoPS was related to increases in social responsibility and declines in emotion problems. Mediation analyses also showed that paths from integration to CUoPS to subsequent social responsibility and emotional problems were significant (see Table S4). This suggests that

teachers' integration of program strategies may be connected to treatment outcomes through CUoPS.

Contrary to previous research with the WITS Programs, paths from the implementation model to additional treatment outcomes (prosocial leadership, victimization, and aggression) were not significant. It is not known whether this is the result of low power to detect effects in our complex implementation models or whether these treatment outcomes take more time to develop. Longitudinal research with more schools is needed to unravel these possibilities. We also combined teacher and parent reports of children's prosocial leadership and aggression. These are, as is typical, not highly correlated. Reviews of past research (De Los Reyes et al. 2015) importantly suggest that children's behaviors at school and home may differ and may be differentially affected by mental health treatments; however, analyses of these potential differences are beyond the scope of this study. Prosocial leadership may also be a later onset outcome of WITS Programs as it is mainly targeted in developmentally appropriate activities for children in grades 4 to 6.

Limitations

The indices created for this study are face valid, and closely connected to core WITS Programs activities. Only children in schools using the program are included in these analyses. Hence, the findings of this study are limited to the implementation of the WITS Programs in rural Canadian elementary schools. Although the SES of the schools and children's families were diverse, most were Caucasian. All data were collected in program schools that had agreed to use the WITS Programs as part of an effectiveness evaluation, so implementation was high. Although our use of Bayesian analysis allows us to examine a cross-classified theoretical model that accounted for the change in teachers, fit statistics are not provided for our models because appropriate summary statistic within Bayesian CFA "protecting against an undesirably high sensitivity to detect negligible differences within large samples" are not currently available (Hoofs et al. 2018, p. 560).

Multiple factors that can affect implementation (Greenberg et al. 2005; Berkel et al. 2011; Proctor et al. 2010) are not examined here (e.g., staff turnover, school levels of poverty, readiness to implement a program). Variability in school configurations may also have affected the responses of older program students. While some schools were kindergarten to grade 7, others were kindergarten to grade 5, and so children moved to middle schools that were not implementing WITS Programs. In addition, most schools were already using additional social emotional learning programs (e.g., *Second Step, Restitution)*. Evidence for the additive effects of implementing the WITS Programs strategies beyond these other programs could not be tested. Additionally, our study focused primarily on between-person differences over time (while controlling for dependencies related to teachers) and did not assess the effects of different levels of analysis (e.g., between-classroom or between-school).

Conclusions

Our findings point to the potential advantages of incorporating and assessing user activities and processes that impact implementation fidelity. Users' efforts to integrate program strategies into everyday practices and children exhibit use of program strategies were related in feedback loops to implementation fidelity in the current research. Findings of this study also demonstrate that fidelity and integration into daily activities and CUoPS have important effects on each other and on child treatment outcomes. Developing concise, inexpensive measures to help schools track the fidelity and to illuminate the activities that they use to sustain fidelity over time may help them sustain their investments in program implementation.

Funding This research was funded by the Public Health Agency of Canada's (PHAC) Innovation Strategy; Entitled: *Taking Action to Reduce Health Inequalities in Canada*. The findings and conclusions of this research are those of the authors and do not necessarily represent the position of the PHAC.

Compliance with Ethical Standards

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the University of Victoria Human Subjects Review Board.

Informed Consent Informed, written consent was obtained from all adults involved in the study. A parent also gave consent for the children to be involved in the research.

Conflict of Interest Dr. Leadbeater is a co-developer and long term evaluator of the WITS Programs. The other authors declare that they have no conflicts of interest.

References

Andreou, T. E., McIntosh, K., Ross, S. W., & Kahn, J. D. (2015). Critical incidents in sustaining school-wide positive behavioral interventions and supports. *The Journal of Special Education, 49*, 157–167. https://doi.org/10.1177/0022466914554298o.

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science, 12*, 23–33. https://doi.org/10.1007/s11121-010-0186-1.

Berkel, C., Mauricio, A. M., Sandler, I. N., Wolchik, S. A., Gallo, C. G., & Brown, C. H. (2017). The cascading effects of multiple dimensions of implementation on program outcomes: A test of a theoretical model. *Prevention Science*. https://doi.org/10.1007/s11121-017-0855-4.

Caldwell, C. B., & Pianta, R. C. (1991). A measure of young children's problem and competence behaviors: The Early School Behavior Scale. *Journal of Psychoeducational Assessment, 9*, 32–44.

Crick, N. R., & Grotpeter, J. K. (1996). Children's treatment by peers: Victims of relational and overt aggression. *Development and Psychopathology, 8*, 367–380.

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*, 858–900. https://doi.org/10.1037/a0038498.

Desjardins, T., Yeung Thompson, R. S., Sukhawathanakul, P., Leadbeater, B. J., & MacDonald, S. W. (2013). Factor structure of the Social Experience Questionnaire across time, sex, and grade among early elementary school children. *Psychological Assessment, 25*, 1058–1068.

Dusenbury, L., Brannigan, R., Hansen, W. B., Walsh, J., & Falco, M. (2004). Quality of implementation: Developing measures crucial to understanding the diffusion of preventive interventions. *Health Education Research, 20*, 308–313.

Forman, S. G., Shapiro, E. S., Codding, R. S., Gonzales, J. E., Reddy, L. A., Rosenfield, S. A., et al. (2013). Implementation science and school psychology. *School Psychology Quarterly, 28*, 77–100. https://doi.org/10.1037/spq0000019.

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation, 17*, 1–13.

Giesbrecht, G. F., Leadbeater, B. J., & MacDonald, S. S. (2011). Child and context characteristics in trajectories of physical and relational victimization among early elementary school children. *Development and Psychopathology, 23*, 239–252. https://doi.org/10.1017/S0954579410000763.

Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A., & Zins, J. E. (2005). The study of implementation in school-based preventive interventions: Theory, research, and practice. Promotion of mental health and prevention of mental and behavioral disorders 2005 Series V3.

Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology, 33*, 665–679. https://doi.org/10.1007/s10802-005-7646-2.

Hansen, W. B. (2014). Measuring fidelity. In Z. Sloboda & H. Petras (Eds.), *Defining prevention science* (pp. 335–359). New York: NY. Springer.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science, 24*, 1918–1927. https://doi.org/10.1177/0956797613480187.

Hirschstein, M. K., Edstrom, L. V. S., Frey, K. S., Snell, J. L., & MacKenzie, E. P. (2007). Walking the talk in bullying prevention: Teacher implementation variables related to initial impact of the Steps to Respect program. *School Psychology Review, 36*, 3–21.

Hoglund, W. L. G., Hosan, N. E., & Leadbeater, B. J. (2012). Using your WITS: A 6-year follow-up of a peer victimization prevention program. *School Psychology Review, 41*, 193–214.

Hoofs, H., van de Schoot, R., Jansen, N. H., & Kant, I. (2018). Evaluating model fit in Bayesian confirmatory factor analysis with large samples: Simulation study introducing the BRMSEA. *Educational and Psychological Measurement, 78*, 537–568. https://doi.org/10.1177/0013164417709314.

Indig, D., Lee, K., Grunseit, A., Milat, A., & Bauman, A. (2017). Pathways for scaling up public health interventions. *BMC Public Health, 18*, 68. https://doi.org/10.1186/s12889-017-4572-5.

Leadbeater, B. J., & Sukhawathanakul, P. (2011). Multi-component programs for reducing peer victimization in early elementary school: A longitudinal evaluation of the WITS® Primary program. *Journal of Community Psychology, 39*, 606–620.

Leadbeater, B. J., Gladstone, E. J., & Sukhawathanakul, P. (2015). Planning for sustainability of an evidence-based mental health promotion program in Canadian elementary schools. *American Journal of Community Psychology, 56*, 120–133. https://doi.org/10.1007/s10464-015-9737-8.

Leadbeater, B. J., Thompson, K., & Sukhawathanakul, P. (2016). Enhancing social responsibility and prosocial leadership to prevent aggression, peer victimization, and emotional problems in elementary school children. *American Journal of Community Psychology, 58*, 365–376. https://doi.org/10.1002/ajcp.12092.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford Press.

Low, S., Smolkowski, K., & Cook, C. (2016). What constitutes high-integration implementation of SEL programs? A latent class analysis of second step® implementation. *Prevention Science, 17*, 981–991. https://doi.org/10.1007/s11121-016-0670-3.

McLachlan, G. J., Krishnan, T., & Ng, S. K. (2004). The EM algorithm (Paper No. 2004, 24). Berlin, Germany: Humboldt-Universität Berlin, Center for Applied Statistics and Economics (CASE).

Muthén, L. K., & Muthén, B. O. (1999-2012). *Mplus user's guide*. Los Angeles: Muthén and Muthén.

Ogden, T., & Fixsen, D. L. (2014). Implementation science: A brief overview and a look ahead. *Zeitschrift für Psychologie, 222*, 4–11. https://doi.org/10.1027/2151-2604/a000160.

Owens, J. S., Lyon, A. R., Brandt, N. E., Warner, C. M., Nadeem, E., Spiel, C., & Wagner, M. (2014). Implementation science in school mental health: Key constructs in a developing research agenda. *School Mental Health, 6*, 99–111. https://doi.org/10.1007/s12310-013-9115-3.

Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health and Mental Health Services Research, 36*(1), 24–34.

Proctor, E. K., Powell, B. J., & McMillen, C. J. (2013). Implementation strategies: Recommendations for specifying and reporting. *Implementation Science, 8*, 139. https://doi.org/10.1186/1748-5908-8-139.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., & Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76.

Reynolds, C. R., & Kamphaus, R. W. (2004). *BASC-2: Behavior assessment system for children*. Minneapolis: NCS Pearson.

Smith, E. P., Osgood, D. W., Oh, Y., & Caldwell, L. C. (2018). Promoting afterschool quality and positive youth development: Cluster randomized trial of the Pax Good Behavior game. *Prevention Science, 19*, 159–173. https://doi.org/10.1007/s11121-017-0820-2.

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. https://doi.org/10.1111/cdev.12864.

Wanless, S. B., & Domitrovich, C. E. (2015). Readiness to implement school-based social-emotional learning interventions: Using research on factors related to implementation to maximize integration. *Prevention Science, 16*, 1037–1043.