Quality of implementation of a school mental health initiative and changes over time in students' social and emotional competencies

Helen Askell-Williams [a], Katherine L. Dix [a], Michael J. Lawson [a] & Phillip T. Slee [a]

[a] School of Education, Flinders University, Adelaide, South Australia
Published online: 12 Jul 2012.

PLEASE SCROLL DOWN FOR ARTICLE

Routledge
Taylor & Francis Group

Quality of implementation of a school mental health initiative and changes over time in students' social and emotional competencies

Helen Askell-Williams*, Katherine L. Dix, Michael J. Lawson and Phillip T. Slee

School of Education, Flinders University, Adelaide, South Australia

This paper reports the theoretical conceptualisation, statistical development, and application of an Implementation Index to evaluate the quality of implementation of the KidsMatter Primary school mental health initiative in Australia. Questionnaires were received from the parents and teachers of almost 5000 students, and also from KidsMatter project officers. A conceptual framework of fidelity, dosage, and delivery guided the selection of questionnaire items to create the Implementation Index, which was refined using Latent Class Analysis. Schools' scores on the Index were classified into high, average, and low implementation categories. Profiles of high- and low-implementing schools provided insights into the characteristics of successful and less successful implementation. Next, hierarchical linear modelling showed that children's social and emotional competencies significantly improved over time in average- and high-implementing schools, but not in low-implementing schools. The Implementation Index can inform areas for attention in health promotion initiatives and can provide a framework for future evaluations.

Keywords: mental health promotion; evaluation; implementation quality; latent class analysis; hierarchical linear modelling

Introduction

Student wellbeing is of central concern for parents[1], teachers, and governments. In recent times, several organisations have undertaken major school-based initiatives to address students' wellbeing. Schools have ready-made populations of students that can be targeted for general, as well as specific, mental health promotion initiatives (Collaborative for Academic Social and Emotional Learning [CASEL], 2008; Domitrovich et al., 2008; Giesen, Searle, & Sawyer, 2007; Mukoma & Flisher, 2004; World Health Organiziation [WHO] Europe, 2006). The focus of such initiatives has moved, in accordance with World Health Organization recommendations, towards a "settings" approach, which is reflected in the concept of the health promoting school (Mukoma & Flisher, 2004; WHO and Health and Welfare Canada, 1986). Such initiatives include the Social and Emotional Aspects of Learning (SEAL) program in the United Kingdom (Department for Children, Schools and Families, 2009) and initiatives by the United States-based Collaborative for Academic, Social and

---

*Corresponding author. Email: helen.askell-williams@flinders.edu.au

Emotional Learning (CASEL, 2011a), the European Centre for Educational Resilience and Socio-Emotional Health (EDRES, 2009), and the European Network for Socio-Emotional Competence in Children (ENSEC, 2009). Initiatives in Australia include MindMatters Secondary (MindMatters, n.d.), KidsMatter Primary (KidsMatter, 2009), and KidsMatter Early Childhood (KMEC, n.d.).

In order to convince stakeholders that health promotion initiatives are worthwhile investments, there is a need for strong evidence that the initiatives do make a difference to school environments and student wellbeing (Biglan, Mrazek, Carnine, & Flay, 2003; Mukoma & Flisher, 2004; Rootman, 2001). Colquhoun (2005) argued for realistic evaluation (see Pawson & Tilley, 1997) to answer questions such as, Does health promotion in school settings work?, Under what circumstances does health promotion in schools work?, How does health promotion delivered through school programs work?, and, For whom do such programs work?

It becomes apparent from these questions that evaluations can be of different types. An *outcome* evaluation focuses on whether an initiative achieved its stated aims. A *resources* evaluation examines whether the structural supports and materials used as the core of the initiative (e.g., curriculum materials and professional development programs) were of good quality. A third type, *process* evaluation, focuses on the way the initiative was enacted. These three evaluation types are interrelated (Payne, 2009). For example, good quality curriculum resources could be accompanied by poor professional learning programs and subsequent poor quality delivery processes, leading to poor outcomes. Or, in an alternative scenario, poor quality resources could be compensated for by creative and skilled delivery, with positive outcomes being achieved in spite of material deficiencies. The findings of all three evaluation types, outcomes, resources, and processes, can contribute to policy decisions about ways to proceed with the subsequent roll-out of health promotion initiatives.

Evidence from *outcome* evaluations is growing. For example, a recent review by Durlak, Weissberg, Dymnicki, Taylor, and Schellinger (2011) indicated that rigorous assessments of outcomes of mental health promotion initiatives in schools demonstrate that such programs can have an impact upon students' social and emotional skills and academic performance. The *processes* of implementation are less well documented but are receiving increased attention (Elias, Zins, Graczyk, & Weissberg, 2003). In particular, identifying relationships *between* processes of implementation and outcomes has been identified as needing an increased research focus (Conduct Problems Prevention Research Group, 2010; Domitrovich & Greenberg, 2000).

The present paper attends to these latter two issues. First, we report the use of latent class analysis (LCA) to develop and apply an Implementation Index to inform the evaluation of implementation processes during the KidsMatter Primary school mental health initiative in Australia. Second, we report the use of hierarchical linear modelling (HLM) to identify relationships between schools categorised as high or low implementers on the Implementation Index, and changes over time in students' scores on a social and emotional competencies scale.

Interventions and evaluations in school-based settings

Schools are complex organisations that pose significant challenges for the delivery and evaluation of health promotion initiatives (Barry & Jenkins, 2007; Clift &

Jensen, 2005; Payne, 2009). Recently, Resnick (2010) drew attention to how the structural affordances and constraints of educational organisations facilitate the success or failure of educational initiatives. Even within a cluster of settings that may be structurally alike in some ways (such as schools within the same educational system), conditions can vary widely. For example, Askell-Williams, Lawson, and Slee (2009) discussed a range of personal and social conditions, such as students' and teachers' background knowledge, existing programs, availability of resources, and leadership commitment to the aims of the initiatives, that vary across schools prior to the introduction of any new initiative. Such pre-existing conditions can be expected to interact with the components of an initiative in either supportive or non-supportive ways. Similarly, Jaycox et al. (2006) argued that the success of a program may rely upon whether sufficient background work has been done to match the intervention program to the needs of the target population, and whether that population is effectively engaged with the implementation of the intervention. Lee et al. (2008) also expressed concerns about rolling out evidence-based prevention programs into complicated real world conditions with limited controls over implementation processes. In such settings, key program components might be modified or deleted, and inconsistencies in program delivery could develop. For example, programs with multiple intervention components (e.g., child, parent, and school), and extended time lines, place substantial demands on local implementers (such as teachers), who may have little or no previous experience in delivering mental health interventions. Lee et al. warned that, over time, complex interventions may be simplified at the expense of key program objectives and strategies, with consequent impacts on program outcomes.

The designs used for many contemporary, settings-based health promotion initiatives have been described by Weare (2010) as loose enabling frameworks. Such intervention designs may be based on a common approach at a broad conceptual level, but may not explicitly specify components for action in each setting. Rather, the model of intervention may be for each setting to determine which programs, from a potentially broad range of related offerings, best suit their needs. With particular reference to school-based health promotion initiatives, Weare and Nind (2011) proposed that "something is going wrong either with the evaluation of many agency-led European and Australian whole school approaches, or with the approaches themselves" (p. 62). Weare and Nind explained this further as follows:

> Some whole school approaches ... are failing to show impact (Wilson and Lipsey, 2007; Durlak et al., 2011). Authors attribute this to a lack of consistent, rigorous and faithful implementation which is causing these approaches to become too diluted and thereby to lack impact. The European and Australian style and the type of whole school approaches it generates tend to promote 'bottom up' principles such as empowerment, autonomy, democracy, and local adaptability and ownership (WHO, 1997). [Such programs] have produced a wealth of well-planned materials, guidelines and advice, but are also deliberately non-prescriptive and principles based. This flexible and non-prescriptive style is echoed in wider approaches to mental health across Europe and Australia, which emphasise the need for end-user involvement and the lay voice. This approach contrasts with the US style of more top-down, manualised approaches, with scripts, prescriptive training, and a strict requirement for programme fidelity. (p. i66)

By way of example, in a randomised controlled study to evaluate health promoting schools in Australia, Mitchell, Palmer, Booth, and Davies (2000 p. 245) concluded that "significant differences in health-related policy and practice would

only be seen if schools focused on the same issues'', which the schools in the study did not do. Another example was provided by Melhuish et al. (2007), who reported the need for innovative approaches to the evaluation the Sure Start Local Programmes, where local autonomy, combined with lack of specification of how the program aims were to be achieved, meant that there was no specified model against which fidelity could be formally evaluated. Similarly, Humphrey, Lendrum, and Wigelsworth (2010) posited that the very small effect sizes for desired outcomes in SEAL appeared to be related to lack of structure and consistency in social and emotional education programs, un-monitored delivery and an inadequate level of human and financial resources, related to its design being,

> essentially what individual schools make of it rather than being a single, consistently definable entity. It was conceptualised in this manner to avoid the lack of ownership and sustainability that might be associated with the more ''top-down'', prescribed approach that is taken in the USA. (Humphrey et al., 2010 p. 7)

Concerns about these ''common framework-variable program'' designs are significant. However, because school-based interventions constitute such a major proportion of contemporary mental-health promotion initiatives, it is important to accommodate the fact that such initiatives often involve quite distinct intervention designs. Although it might be optimal to consider only interventions that have been conducted in replicated randomised control trials (see Bellg et al., 2004; Weare & Nind, 2011), Rootman (2001) argued that the field of health promotion may not always lend itself to such ''high-level'' efficacy assessments to measure intervention successes. Interventions that have proven efficacious under well-controlled conditions may have high internal validity but low external validity when delivered in variable contexts such as schools (Kam, Greenberg, & Walls, 2003).

Given the substantial level of financial resources and professional time invested in interventions such as SEAL or KidsMatter, it is important that key factors associated with the intervention effects be identified. This will require innovative approaches to evaluation design. One such approach is to examine the influence of implementation processes on the effect of an intervention.

Process evaluations of school-based interventions

This line of reasoning about the need to attend to implementation processes was adopted by Domitrovich et al. (2008) when they considered why some programs that are successful in efficacy trials have been shown subsequently to be less successful in real school settings. They suggested that a key factor might be the quality of the processes of implementation, which could be predicted to be associated with the outcome of the intervention. Melde, Esbensen, and Tusinski (2006) argued that process evaluation can document observed responses to, for example, the frequency and quantity of delivery of components of the intervention, including whether such dosage reflects the program intentions. A significant advantage of process evaluation, especially with respect to programs that appear to fail, is that it can inform analyses of whether a failure is due to poor program design or to poor implementation of a (potentially well-designed) program. Similarly, process evaluation permits analyses of successful program characteristics: Such information is vital for efficient transfer of programs to new contexts.

Frameworks for process evaluations

What should be considered in a process evaluation? Lee et al. (2008), drawing from earlier work of Dane and Schneider (1998), described a multidimensional construct of implementation fidelity, consisting of the number and length of services provided (exposure), the degree to which program strategies conformed to the manual (adherence), and how well implementers delivered the program (quality of delivery, including implementer enthusiasm, comportment, and preparedness). A similar approach to process evaluation was also developed by Domitrovich et al. (2008), who identified the following three dimensions of implementation quality:

- Fidelity: The degree to which an intervention and its support system are conducted as planned.
- Dosage: The frequency, duration, and timing of both the interventions (such as numbers of lessons delivered) and their support systems (such as number of contacts between project officers and school principals).
- Delivery: The professional expertise of the delivery agents (e.g., teachers, support personnel, and external agents), including issues such as sensitivity, engagement, and responsiveness to the goals of the program and to the recipients of the program in terms of productive teacher–learner relationships.

It is these dimensions of fidelity, dosage, and delivery that we used to frame the evaluation of the processes of implementation in the KidsMatter initiative trial and that are reported in this paper.

The above three dimensions have been considered to varying degrees in past evaluations. For example, Kam et al. (2003) identified two factors that contributed to intervention success, namely, (a) adequate support from school principals and (b) high degree of classroom implementation by teachers. In a wide-ranging treatment, Shek, Sun, and Kan (2009) analysed classroom-based observations of the Positive Adolescent Training through Holistic Social Programmes (PATHS) in Hong Kong. Shek et al. identified six factors that represented good implementation quality. The first factor was program adherence and achievement of program objectives. The next four factors were related to instructors' program delivery skills and preparation, including carrying out the program in an interactive manner, giving positive feedback and support to students, providing opportunities for student reflection, and good lesson preparation. The sixth factor was student interest in attending the lessons. The factors identified by Shek et al. and Kam et al. could be used as indicators of processes of delivery, and fidelity, respectively.

In a study of a different PATHS (Promoting Alternative Thinking Strategies) program, Greenberg and colleagues (Conduct Problems Prevention Research Group, 2010) collected dosage information from weekly teacher reports about the PATHS lessons they had presented and fidelity information from observers' assessments. Four indicators of fidelity were employed: (a) quality of teaching of PATHS concepts, (b) teachers' modelling of PATHS concepts throughout the school day, (c) quality of classroom management (during PATHS lessons), and (d) openness of teachers to consultation with the educational consultant. These observations appear to provide broad assessments of fidelity to the overall aims of

the program and might also be relevant indicators of the quality of delivery of the program.

Van Der Heyden, Witt, and Gilbertson (2007) drew attention to aspects of intervention support, arguing that interventions can decrease the need for special education services when implemented with high integrity by a paid research associate. However, these authors raised questions about whether such dedicated support is sustainable in regular settings. They posited the more likely scenario that, in the longer term, teachers (alone) deliver the interventions with consequent losses of program fidelity. Similarly, Melde et al. (2006) noted that school-based prevention or intervention programs often failed to implement interventions in ways consistent with program expectations. Melde et al. reported an evaluation of a classroom intervention in which fidelity was compromised by typical deviations from the prescribed program, such as omission of the introduction or conclusion, skipping steps in the lesson, or becoming so engrossed in some steps that lesson time ran out. Indeed, in the Melde et al. report, time management was an issue for the 11 classes observed. The observed events in the Melde et al. study could result in reduced program effects and cause program providers to overestimate the fidelity of their own implementation practices. Similarly, Humphrey et al. (2010) reported some of the observed variations in the processes of implementation of SEAL in the UK, including variations in the classifications of staff who delivered the modules of instruction, how long and how often the instruction occurred, and how much status the instruction was afforded by staff. And Hallam's (2009) report from an analysis of the SEAL evaluation data demonstrated how interactions between implementation processes and outcomes might occur. Hallam indicated that SEAL had led to outcomes such as increase in staff' understandings of the social and emotional aspects of learning and increases in staff' understandings about their pupils. This had changed staff behaviour, enhanced their confidence in their interactions with pupils, and led them to approach behaviour incidents in a more thoughtful way. It is arguable that these outcomes related to staff' capabilities would have interacted with the processes of the staff' delivery of SEAL and, as such, address the delivery dimension suggested in the Domitrovich et al. (2008) framework.

Evaluating implementation

From the above review, it can be seen that quality of processes of implementation is an issue of ongoing concern. However, evaluations of implementation quality have, mostly, not focussed on the range of potential influences encompassed by the dimensions of the Domitovich et al. (2008) framework. Indeed, Domitrovich and Greenberg (2000), Melde et al. (2006), Lee et al. (2008), Payne (2009), and Greenberg (2011) have all identified gaps in the assessment and documentation of the processes of implementation, with research reports consistently failing to make use of information related to processes of program delivery. Furthermore, Greenberg argued that the field of prevention science, to which mental health promotion belongs, requires new and innovative processes and tools for measuring the effectiveness of interventions. Thus, one purpose of the present paper is to report the design of an Implementation Index that provides measurements of the three dimensions of fidelity, dosage, and delivery in the implementation of the KidsMatter intervention.

A second purpose of this paper addresses another issue identified by Domitrovich and Greenberg (2000), namely, that relatively few studies report relationships between the quality of processes of implementation and desired program outcomes for students. For example, Domitrovich and Greenberg reported that, in 34 prevention programs, only 21% examined whether any aspect of the quality of the intervention was associated with desired outcomes. One example where a fidelity assessment *was* matched to outcomes was provided by Oshima, Cho, and Takahashi (2004), who used fidelity scores to divide subjects into three groups (low, medium, and high fidelity). The authors then compared the outcomes between participants in the low-fidelity and high-fidelity implementation groups. Another example of matching implementation quality to outcomes was provided by Melhuish et al. (2007), who, in their evaluation of the Sure Start Local Programmes, reported that programs scoring higher on 18 purpose-designed implementation proficiency indicators also scored higher on measures of program impact on child and parenting outcomes.

Recognising that relationships do exist between processes and outcomes, Kam et al. (2003) proposed that reports about "effectiveness" (processes) need to accompany reports about "efficacy" (outcomes). Such an approach recognises the difficulties faced in synthesising process, output, and outcome findings, while attending to different aspects and levels of activities within and between different types of schools (Mukoma & Flisher, 2004). Thus, to address the second aim of the present paper, we report associations between schools' scores on the purpose-designed Implementation Index and an outcome measure of children's social and emotional competencies.

The objectives of this paper are guided by the following research questions:

(1) Can an Implementation Index based upon a three-dimensional framework of fidelity, dosage, and delivery reliably differentiate between high-implementing and low-implementing KidsMatter schools?
(2) What are the relationships between high-implementing and low-implementing KidsMatter schools and changes over time in students' social and emotional competencies?

The context in which these research questions are addressed is the KidsMatter Primary school mental health promotion, prevention, and early intervention initiative (KidsMatter), which was piloted in 101 Australian primary schools in 2007–2008.

The KidsMatter Primary Schools Mental Health Initiative

KidsMatter was founded on a social-ecological approach that recognises the influences of parenting actions and family environments, school contexts, and the psychological world of each child, on the development of children's positive mental health (Graetz et al., 2008; Slee et al., 2009). Expressions of interest to join the two-year KidsMatter pilot were sought from all 7739 Australian primary and middle schools. From the resulting 260 expressions of interest, a stratified sampling procedure was used to select 101 schools.[2] The sampling procedure ensured the participation of schools located in metropolitan, rural, and remote areas, the inclusion of state, independent, and Catholic schools, and encompassed a range of socioeconomic areas.

The aims[3] of KidsMatter were to:

- improve the mental health and well-being of primary school students,
- reduce mental health difficulties amongst students, and
- achieve greater support for students experiencing mental health difficulties.

KidsMatter provided schools with a common conceptual framework, a prescribed implementation process, and resources such as program and curriculum guides, web- and paper-based information, staff professional learning programs, and KidsMatter project officer support. The KidsMatter approach of providing a common conceptual framework, but leaving freedom for school sites to select programs suitable for their own contexts, was intentionally designed to respond to variations in school settings located across wide physical, social, and cultural geographies. This approach was similar to that used in SEAL (Humphrey et al., 2010) and the Sure Start Local Programmes (Melhuish et al., 2007) and was one of the democratic approaches to intervention design specifically referred to by Weare and Nind (2011). In the KidsMatter model, State-based KidsMatter project officers supported staff from each participating site to identify which programs were needed to improve existing services, and which activities needed to be adopted or created to meet identified needs, without direct specification of how existing school programs were to be changed or exactly what was to be delivered.

The KidsMatter conceptual framework consisted of a whole-school approach with four nominated components for intervention, namely, (1) developing a positive school community, (2) providing regular social and emotional education for students, (3) providing parenting support and education, and (4) early intervention for students experiencing, or at risk of, mental health difficulties. In addition, KidsMatter prescribed that schools should follow a 7-Step implementation process related to focussing attention on the four components, as follows:

Step 1. Define the issues by writing a summary statement to describe the school's current situation related to each component
Step 2. Set goals based on each summary statement
Step 3. Identify any concerns in achieving the goals
Step 4. Develop a broad range of options/strategies to address concerns and achieve goals
Step 5. Evaluate the feasibility of each option/strategy
Step 6. Formalise the component plan
Step 7. Implement the plan and review

KidsMatter schools were provided with a manual; however, it was not a manual of programs to be implemented. Rather, the manual prescribed processes for schools to follow in order to address each component according to school needs. For example, the manual provided advice about conducting a school audit to identify areas of strength and weakness, and prescribed that social and emotional education should be delivered to all students regularly, and at least once per week. Schools were free to select their own resources, such as curricula for social and emotional education, and were provided with an annotated guide to a wide range of social-emotional programs that could inform their curriculum choices. Implementation support and staff professional development was coordinated, and often personally

provided, by KidsMatter project officers, who were each responsible for approximately ten schools.

Evaluating KidsMatter schools' attention to the four component framework and 7-Step implementation process was central to an extensive collection of quantitative and qualitative data across the 2 years of the pilot initiative. Some of these data underpin the analysis and discussion presented in this paper.

Method

Ethics approvals

We received ethics approvals from our University's Social and Behavioural Research Ethics Committee, and from all school, jurisdiction, and departmental bodies in all Australian states and Territories.

Questionnaires from multiple informants

Following an extensive review of the literature, we created questionnaires for parents, teachers and KidsMatter project officers. The questionnaire items were piloted (with participants not involved in KidsMatter) for face and content validity. Where scales were created, the scale items were subjected to exploratory and confirmatory factor analysis and tested for item reliability.[4] The use of responses from teachers and parents within each school, as well as responses from non-school-based (external) KidsMatter project officers who were responsible for clusters of schools, enabled us to triangulate a range of informants' perspectives about concepts of interest.

We delivered the questionnaires on four occasions over 2 school years to the teachers and parents of a stratified sample of over 7000 students, resulting from the selection of up to 76 students per school, with approximately equal numbers of boys and girls in the KidsMatter targeted age group of 10 years old (backfilled with students from different age groups in smaller schools, in order to achieve up to 76 students per school[5]). The questionnaires covered two main areas: (1) engagement with, and implementation of, KidsMatter, and (2) influences of KidsMatter on schools, teachers, parents, and students. Most items on the questionnaires required responses on a 7-point Likert rating scale, ranging from *strongly disagree* (score 1) to *strongly agree* (score 7). In particular, the teacher questionnaires included a scale to measure each selected child's social and emotional competencies (detailed later in this paper).

Eight state-based KidsMatter project officers, who were responsible for facilitating the intervention in schools, also completed questionnaires about various aspects of the implementation of KidsMatter in each of the schools to which they were assigned. In particular, project officers were able to make relative assessments across schools.

Multiple contexts and occasions

Fifty schools (termed Round 1) commenced KidsMatter in 2007, and the other 50 schools (termed Round 2) commenced in 2008. We collected questionnaire data from all Round 1 and Round 2 schools on four occasions (in the first and last school terms of each of the 2 school years). An overview of the

participants, evaluation design, and the selected outcome measure is presented in Figure 1.

Results

The response rate from parents and teachers at Time 1 was 70% ($n = 4970$) of the original sample, which provided information about students having a mean age of 9.7 years ($SD = 1.6$) and near-equal representation of boys (48%) and girls (52%). Of these initial responses, 76% of parents and teachers continued to submit questionnaires on the remaining three data collection occasions.

Treatment of missing data

Analysis indicated that missing data through non-participation was not random, with underrepresentation of students from Aboriginal or Torres Strait Islander backgrounds, students identified as ''at risk'', and students from culturally and linguistically diverse backgrounds. Natural attrition, student transience, and postal error led to a 24% reduction in the initial response rate by Time 4. Data missing within and between occasions arose from different causes, and thus were treated conservatively by selecting statistical procedures that could accommodate missing data, rather than imputing missing values. Nevertheless, it was necessary to impute some missing data at the aggregated student level for the requirements of hierarchical linear modelling (HLM), which allows for missing data at Level 1 but requires complete data at Levels 2 and 3 (Raudenbush, Bryk, Cheong, & Congdon, 2004). In the HLM, missing data on the aggregated variables were below 5% and were replaced with the local median (see Dix et al., 2010).

The Implementation Index

Using a conceptual framework based on the Domitrovich et al. (2008) dimensions of fidelity, dosage, and delivery, we undertook a review of the content of items in the parent, teacher, and project officer questionnaires. Given the broad approach of the
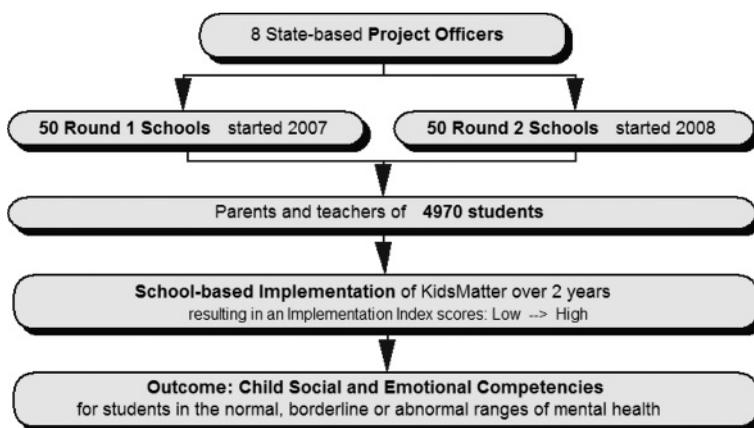

Figure 1.   Overview of the participants, research design, and outcomes.

KidsMatter initiative, we identified items related to universal features of implementation, such as the prescribed 7-Step implementation process, leadership involvement, and communication with parents/carers. The conceptual framework and sources of data are presented in the rows and columns, respectively, of Table 1.

We reviewed the content validity of the questionnaire items and selected 50 items for potential inclusion in the Implementation Index. We examined these 50 items using Latent Class Analysis (LCA: Mplus 5.2, Muthén & Muthén, 2007) in order to identify items that best discriminated between schools with higher and lower scores on the selected items. Questionnaire items that discriminated poorly between high and low implementation quality in schools were trimmed from the model. We also considered the differences in the ways Round 1 and Round 2 schools' scores were reflected by each item. For example, due to the staged implementation of KidsMatter, Round 1 schools' scores on the Implementation Index could be interpreted as a reflection of sustained implementation during 2 years, while Round 2 schools scores' were a reflection of relatively short implementation (during 1 year). Hence, we first deleted Index items that were poor indicators in both Rounds of implementation. Next, we reviewed items that were not effective discriminators in one Round, but were effective in the other Round, in conjunction with the overall representation and balance of items across multiple informants. The resulting Implementation Index retained 37 items. Goodness-of-fit statistics were adequate for the purposes of item discrimination.

Table 2 provides details of the 37 items that formed the Implementation Index. From Table 2, it can be seen that schools could achieve a possible minimum Index score of 37, indicating the lowest level of implementation, and a maximum Index score of 226, indicating the highest level of implementation. Meanwhile, Figure 2 presents the LCA probability estimates for the 37 items for each of the Round 1 and Round 2 schools, respectively. The vertical axis can be interpreted as the probability of achieving a high implementation score. Hence, low-implementing schools have lower probabilities of achieving high implementation scores on the items. The LCA profiles suggested that the Implementation Index was able to identify schools that differed in terms of implementation quality.

Table 1. The KidsMatter Implementation Index framework.

	Teacher and Parent Views	Project Officer Views
FIDELITY Degree to which an intervention is conducted as planned	7-Step Implementation Process; Delivery of SEL curriculum	7-Step Implementation Process
DOSAGE Specific units of an intervention and resources	Time for planning and implementation; Principal participation; Amount of professional development	Contact with school leadership; Provision of information to parents
DELIVERY Engagement with the process and support responsiveness	Teacher Rating of PD; Parent engagement	Leadership, staff & parent encouragement and involvement

368   *H. Askell-Williams* et al.

Table 2.   Items comprising the KidsMatter Implementation Index.

| No.	Items	Score Range
FIDELITY: Teachers: For the 4 KidsMatter components, our school:		
1	Has defined the issues	1–7
2	Has set goals	1–7
3	Has identified difficulties in achieving our goals	1–7
4	Has developed strategies for achieving our goals	1–7
5	Has evaluated strategies for addressing the four components	1–7
6	Has developed and implemented coherent plans	1–7
7	Has reviewed and adjusted plans	1–7
8	Our school teaches social and emotional skills to students in formally structured sessions that adhere to a program manual	1–7
FIDELITY: Project Officers: For the components they worked on, this school:		
9	Has defined the issues related to the components	1–7
10	Has set goals for the components	1–7
11	Has identified difficulties for achieving goals for the components	1–7
12	Has developed strategies for achieving goals for the components	1–7
13	Has evaluated strategies for addressing the components	1–7
14	Has developed and implemented plans for the components	1–7
15	Has reviewed and adjusted plans for the components	1–7
16	KidsMatter is well implemented in this school	1–7
DOSAGE: Teachers		
17	Teachers attend professional development associated with KidsMatter	1–7
18	The Principal attends most KidsMatter meetings	Y/N
19	On average, how much time in staff meetings is formally allocated to KidsMatter?	1–3
20	On average, how much formal time per week does the Action team allocate to planning & implementing KidsMatter?	1–3
DOSAGE: Project Officers		
21	How many times did the school provide opportunities for parents to meet with each other?	1–7
22	Did the school send newsletters containing information about parenting home to families?	Y/N
23	Did the school send tip sheets containing information about parenting home to families?	Y/N
24	Did the school send KidsMatter Information sheets home to parents?	Y/N
25	Did you have contact with the Deputy Principal?	Y/N
DELIVERY: Parents		
26	I feel positively about KidsMatter	1–7
27	I am encouraged to participate in KidsMatter	1–7
28	I have formed more support networks with other parents since KidsMatter	1–7
29	I have been more involved with the school since KidsMatter	1–7
30	I feel that the school community is more positive since KidsMatter	1–7
DELIVERY: Teachers		
31	In general, the quality of the Professional Development for KidsMatter has been … (poor to excellent)	1–7
DELIVERY: Project Officers		
32	The school leadership encourages staff to become actively involved with KidsMatter	1–7
33	Staff are actively involved with KidsMatter	1–7
34	The school leadership team is actively involved with KidsMatter	1–7
35	Parents in this school are encouraged to participate in KidsMatter	1–7
36	The whole staff are involved in the planning of KidsMatter	1–7
37	The whole staff are involved in the implementation of KidsMatter	1–7
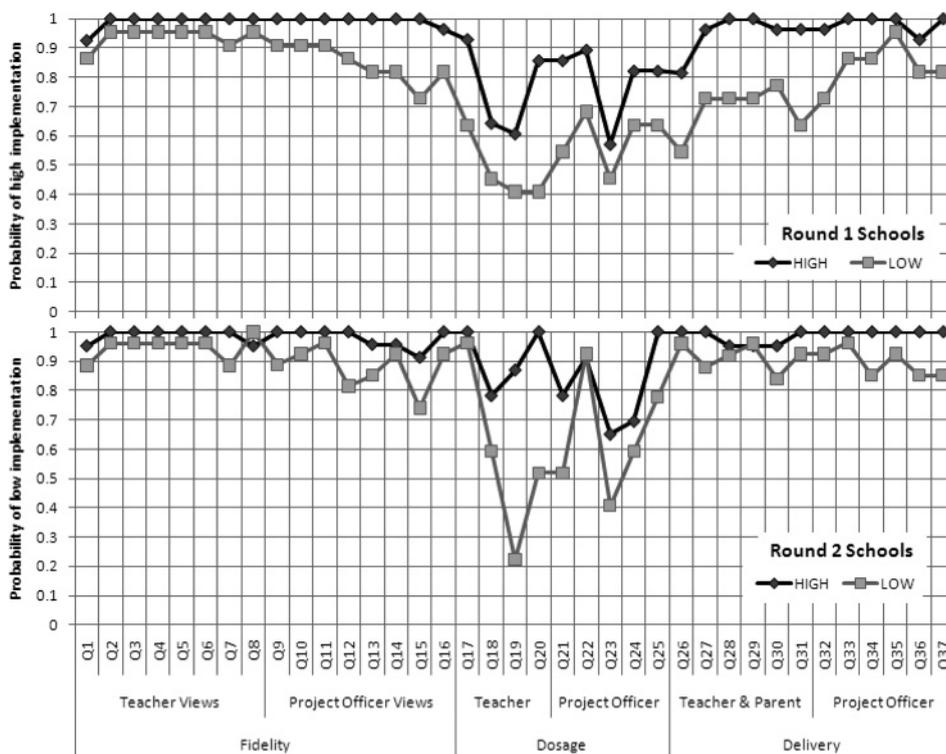
Figure 2.   Latent Class probability estimates of being high or low implementing Round 1 and Round 2 schools, as a means of selecting discriminants.

School profiles on the Implementation Index

While LCA was used to select the Index items based on the probability profiles of schools, the next step in the analysis was to categorise and profile the schools according to their scaled Implementation Index scores. The resulting Index scores, which were calculated by summing responses to each item on a particular occasion, ranged from 56 to 219. Next, we used the visual binning procedure in SPSS to classify schools, according to their score, into low, medium-low, medium-high, and high implementation categories, using cut-points at the mean and ± 1 standard deviations. For simplicity, Figure 3 presents the profiles of the low- and high-implementation groups, calculated using Time 4 data, to illustrate the differences in the standardised mean responses for these schools.

The profiles shown in Figure 3 show substantial differences between low- and high-implementing schools in the areas of Fidelity and Delivery, with less divergence for Dosage. In particular, the Implementation Index indicates that there were observable differences in the extent to which schools progressed on the 7-Step implementation process, both from teachers' ratings (Q1 to Q8) and even more so from project officers' ratings (Q9 to Q16). Teachers' ratings also showed clear differences between groups for the following items:
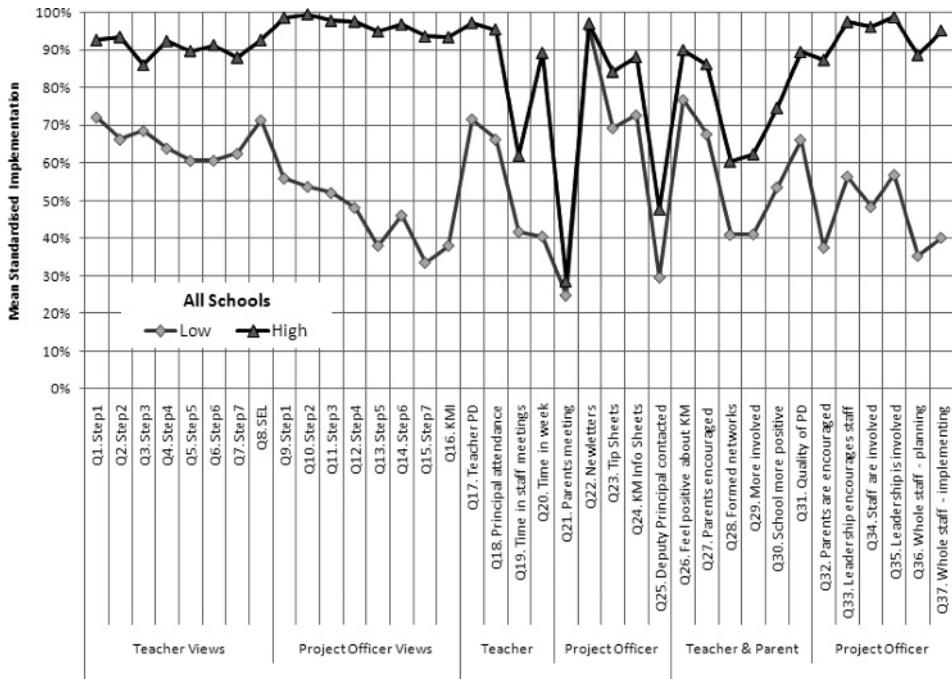
Figure 3.   School profiles on the items of the Implementation Index at Time 4.

On average, how much:

Q19. Time in staff meetings is formally allocated to KidsMatter?

Q20. Formal time per week does the Action team allocate to planning & implementing KidsMatter?

Another major area of difference between high- and low-implementing schools is illustrated to the right of Figure 3, which includes items relating to the involvement and support of parents, staff, and school leadership (Q32 to Q37). Noticeably, these ratings were from the project officers' responses, whose external perspectives of schools showed greater differentiation between high- and low-implementing schools than the internal perspectives of parents and teachers.

Other items of note in the Implementation Index include, Q21: *opportunities for parents to meet with each other*, which was low for all schools, and from parents' perspectives, the two items, Q28: *I have formed more support networks with other parents since KidsMatter* and Q29: *I have been more involved with the school since KidsMatter*, which were relatively low for all schools, but much lower for low-implementing schools. A contrast is provided by Q22*: Send newsletters containing information about parenting home to families*, which was rated highly for all schools.

Validity of the Implementation Index

To consider the predictive validity of the Implementation Index, we compared the performance of the Index under different contexts (Rounds 1 and 2) and on different data collection occasions (Times 1 and 4). The averaged scores of Round 1 and Round 2 schools calculated for Time 1 and Time 4 were standardised to simplify

comparisons. Figure 4 presents the standardised results (expressed as a percentage) for Round 1 and Round 2 schools' scores (55% and 49%, respectively). The difference between the Time 1 scores for Round 1 and Round 2 schools, especially in the Fidelity dimension, is consistent with the delayed KidsMatter initiative start-date (February 2008) for Round 2 schools. Interestingly, Round 1 and Round 2 schools reached similar levels of overall implementation by Time 4 (December 2008; 62% and 65%, respectively). Of particular note are the higher ratings of engagement with the 7-Step process (an aspect of Delivery) in Round 2 schools.

The patterns of findings in these analyses suggest that the Implementation Index was able to discriminate between schools in terms of the components of implementation quality, and that the Index responded appropriately in different contexts and over time.

Relationships between quality of implementation and children's social and emotional competencies

The second research question addressed in this paper is whether the quality of implementation of KidsMatter had a differential effect upon changes over time in students' social and emotional competencies.

Child social and emotional competencies

Improving protective factors for achieving positive mental health was a central focus of KidsMatter. One such protective factor is students' social and emotional competencies (CASEL, 2008; Graetz et al., 2008; WHO, 2007). The questionnaires used in the present study included seven items about students' social and emotional competencies that were based upon the five areas suggested by CASEL (2011b) such as, *This child is happy about his/her relationships with other children* and *This child can manage his/her feelings*. Participating teachers were asked to think about the past month and to rate the extent to which they agreed (7-point scale) that each nominated child had shown each of the named competencies, on average, over the
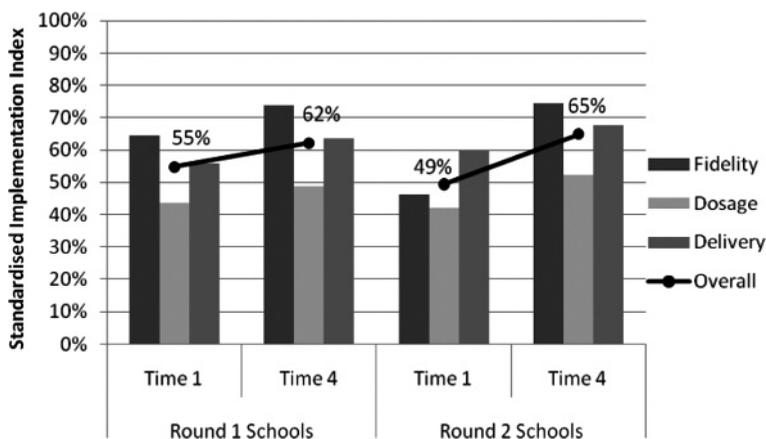

Figure 4. Standardised Implementation Index scores for Round 1 and Round 2 schools at the start (Time 1) and end (Time 4) of the initiative.

month. Factor analysis of the seven items was undertaken using asymptotic distribution-free methods, which was suitable given the large sample size and the skewed distributions of item responses (Garson, 2009; Hox & Bechger, 1998). The analysis confirmed a well-fitting, one-factor model (RMSEA $\leq$ .05; SRMR $\leq$ .06; CFI $\geq$ .95; χ^2/DF $\leq$ 3.0), which we named the student Social and Emotional Competencies (SEC) scale.

Change in students' social and emotional competencies over time

The SEC data, which were collected across the 100 schools on four occasions, provided the opportunity to investigate change over time in social and emotional competencies at the student, and at the school, levels. The nested nature of such data, reflected by the three-level structure of HLM, is presented conceptually in Figure 5. The occasion level (Level 1) contains the within-student outcome variable (SEC) that was rated by teachers for each child on each of the four data collection occasions (TIME). The data structure is best represented by a three-level model. However, for the present analysis, which examines the impact of quality of each school's implementation of KidsMatter upon students' SEC, no between-student variables were entered at Level 2 of the model. The school level (Level 3) contains school variables, including School Round (ROUND) and the Implementation Index groups, identified by the above-mentioned visual-binning procedure (INDEX: low, medium-low, medium-high, high).

The conceptual model presented in Figure 5 was tested in HLM6 (Raudenbush et al., 2004), using $\alpha = .05$ to signify acceptable statistical significance, and a reduction in deviance from the null model to indicate improved model fit. The reliability of the scores and the associated relationship were acceptable ($r > .50$). To facilitate interpretability, all non-dichotomous variables were grand-mean centered. The results of the final three-level modelling are presented in Table 3, showing that the model involved a substantial reduction in deviance (116.81 per parameter estimated), as well as an overall proportion of variance explained of 41%. It should be noted that during the HLM analysis, proportions of variance were relocated amongst the three levels in the process of fitting the model to the data.

The main findings from the HLM presented in Table 3 suggest that SEC improved over the course of the 2 years ($\gamma = 0.07$, $p < .001$), and more so for students in Round 1 schools (2 years of implementation; $\gamma = -0.03$, $p < .05$) that
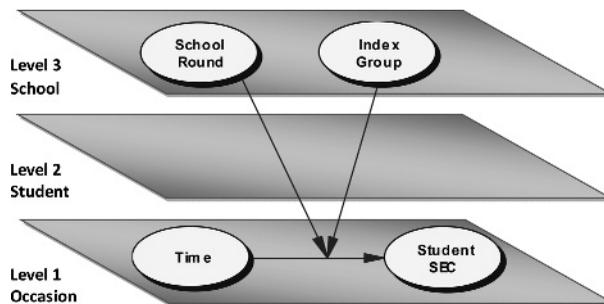

Figure 5.   Conceptual three-level HLM model of child social and emotional competencies to be tested.

were rated as high-implementing schools on the Implementation Index ($\gamma = 0.05$, $p < .001$).

In order to estimate effect sizes, SEC was separately analysed within each Implementation Index group as a function of time. The HLM effect size was determined according to the method described by Dix et al. (2010). Figure 6 shows that the SEC of students in the low-implementing schools did not significantly improve ($\gamma = 0.02$, $p > .05$). However, the SEC of students in the high-implementing schools did significantly improve ($\gamma = 0.09$, $p < .001$), to the extent of a small effect size ($r = .15$).[6] Although not represented in the visual display in Figure 6 (to enable ease of interpretation), it should also be noted that the average-implementation group (based on the combined medium-low and medium-high groups) also yielded a small effect size ($\gamma = 0.07$, $r = .1$, $p < .001$), similar to the high-implementing schools.

Table 3. Final estimations of the three-level HLM student mode.l

Final model: SEC $= \gamma_{000} + \gamma_{100}$TIME $+ \gamma_{101}$ *ROUND*.TIME $+ \gamma_{102}$*INDEX*.TIME $+ u_{00} + u_{10}$TIME $+ e$

Fixed Effect	Coefficient	Std Error	*p* value
INTERCEPT, γ_{000}	5.22	0.04	0.000
TIME, γ_{100}	0.07	0.01	0.000
TIME.ROUND, γ_{101}	-0.03	0.01	0.023
TIME.INDEX, γ_{102}	0.05	0.01	0.000

Deviance Reduced $= 817.70$; Deviance/Additional *df* $= 116.81$; Proportion Variance explained $= 41\%$.

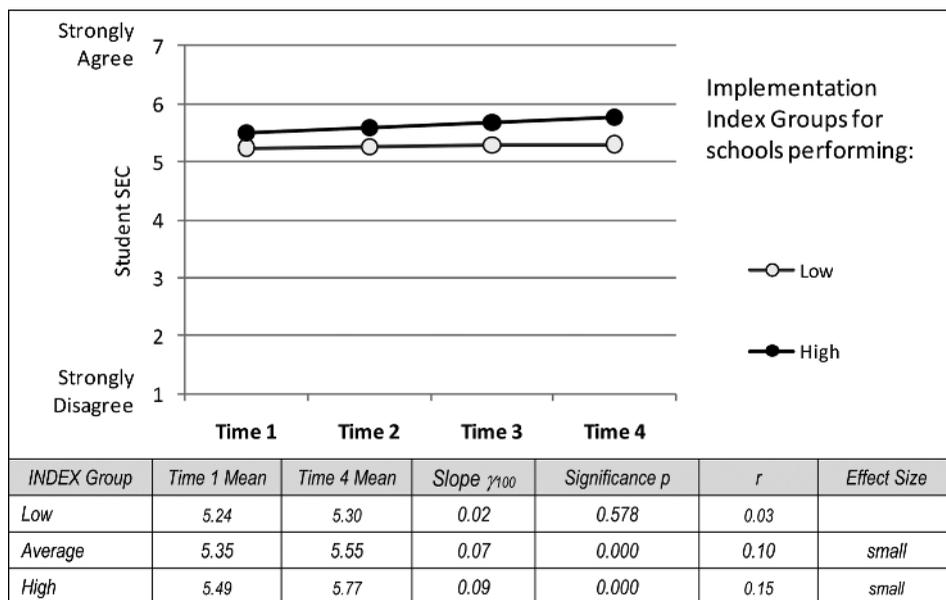| INDEX Group	Time 1 Mean	Time 4 Mean	Slope γ_{100}	Significance *p*	*r*	Effect Size
Low	5.24	5.30	0.02	0.578	0.03	
Average	5.35	5.55	0.07	0.000	0.10	small
High	5.49	5.77	0.09	0.000	0.15	small

Figure 6. Change over time in teacher ratings about student SEC in schools categorised as Low, Average (not shown), or High on the Implementation Index.

Discussion

This paper reports, in the context of the KidsMatter Primary school mental health initiative, the use of latent class analysis (LCA) for the development of an Implementation Index, which was underpinned by a framework of fidelity, dosage, and delivery. The LCA successfully differentiated the probabilities of schools' scores on a range of Implementation Index items, which were selected to represent the three dimensions of the framework. Furthermore, the Implementation Index consistently differentiated between high- and low-implementing schools according to Round, and over four successive data collection events, indicating that the Implementation Index was responsive in different contexts and over time.

Next, the schools' Implementation Index scores were used to categorise schools into high-implementing and low-implementing groups. In particular, differences were observed between high- and low-implementing schools in their adherence to the KidsMatter 7-Step implementation process, indicating a clear direction for required supports in the ongoing roll-out of KidsMatter. Another major area of difference between high- and low-implementing schools was the different level of involvement and support of parents, staff, and school leadership in KidsMatter-related activities. The influence of parent, staff, and leaders' beliefs, supportive capacities, and engagement with initiatives is consistent with literature on the essential components of school reform (e.g., Devaney, O'Brien, Resnik, & Weissberg, 2006; Fullan, 2007; Waks, 2007).

Noticeably, the KidsMatter project officers' external perspectives of schools showed greater differentiation of high- and low-implementing schools than the internal perspectives of parents and teachers. This suggests that the relatively limited experiences of parents and teachers within one school setting (i.e., their own) may have compromised their ability to critically assess the performance of their school, compared to reports from project officers who were able to observe multiple sites. From this finding, evaluators would be well advised to include the perspectives of external respondents, who arguably possess broader outlooks regarding the processes of school-based initiatives.

Importantly, schools' Implementation Index scores at Time 1 illustrate the earlier argument made by Askell-Williams et al. (2009) that initiatives such as KidsMatter may be introduced into settings that are already receptive to providing mental health promotion activities (see Figure 4). Interestingly, Round 1 and Round 2 schools reached similar levels of overall implementation by Time 4, suggesting that Round 2 schools made quicker progress. One possible explanation for this finding is that the later KidsMatter starting date for the Round 2 schools may have enabled those schools to make better advance preparations for its introduction. Examples of advance preparation might include talking to staff about the upcoming initiative, redesigning timetables to accommodate the initiative, and ensuring that resources were delivered to the school in time for the starting date. A second possible explanation is that KidsMatter project officers may have been able to use their experiences gained in working with Round 1 schools to provide more effective support to Round 2 schools. This explanation is reinforced by the higher ratings of engagement with the 7-Step process in Round 2 schools. This finding points to the complex interaction between program duration and program support systems.

Achieving high dosage levels of programmed components was challenging for both Round 1 and Round 2 schools. Specific dosage items that differentiated between high- and low-implementing schools can become the foci of additional

investigation, with a view to the allocation of additional supports such as timetable design, professional learning opportunities, and design and availability of teaching materials.

The findings in this study suggest that use of an Implementation Index to evaluate quality of processes seems particularly useful for health promotion initiatives of the broad conceptual type as described by Weare and Nind (2011) and Melhuish et al. (2007), where evaluation of adherence to tightly prescribed manualised elements is not compatible with the intervention design. Although our findings might be regarded as merely confirming expectations about the impact of implementation quality, the availability of the detailed and theoretically grounded information in the Implementation Index provides an evidence base that has not previously been available. Furthermore, the fine-grained analysis permitted by the Implementation Index responds to Colquhoun's (2005) call for realistic evaluations that recognise complexity and approach history and contexts as key sources of influence on program outputs and effects, not merely as sources of variation to be artificially controlled.

Our second research question enquired whether the quality of implementation of KidsMatter at the school level, as identified by schools' Implementation Index scores, was differentially related to improvements in students' social and emotional competencies. The HLM analysis of change over time (2 years, comprising 4 data collection occasions) showed statistically significant relationships, with small practical effect sizes ($r = .1 - .15$), between the categorisation of schools according to the Implementation Index and improved student social and emotional competencies in average- and high-implementing schools. To state this in more concrete terms, in low-implementing schools, there was a 3.9% shift in children moving from low or average SEC (scores of 5 or less) into high SEC (scores of 6 or more) between Time 1 and 4, which was considered not statistically significant. However, in average- and high-implementing schools, there were 7.8% and 13.1% shifts in children moving from low or average SEC, to high SEC, respectively. These differences were identified in the HLM analysis as small but significant effects. It is particularly noteworthy that the significant improvements in students' SEC were not observed in low-implementing schools, thus showing similarity to the findings of Melhuish et al. (2007).

A final note concerns the practical importance of the small effects found in this study. Rosenthal and DiMatteo (2001) asked whether r was ever too small to matter, giving the example of the effect size r of .034 translating into the important effect of 34 out of every 1000 people being saved from a heart attack if they used low dose aspirin on a regular basis. Rosenthal and DiMatteo presented a table of meta-analytic studies that demonstrated important effects in medicine and psychology that ranged from r's of .09 to .62. Hutchinson and Wilson (2011) noted that small effect sizes are typical in interventions for public health problems, which emphasises the difficulties associated with altering health behaviours. Similarly, Grissom and Kim (2012) stated, "The designations of small, medium and large effect sizes do not necessarily correspond to the degree of practical or theoretical significance of an effect" (p. 130), arguing that small but consistent effects can have considerable impact over time. From the perspective of economic benefit, Foxcroft and Tsertsvadze (2011) and Caulkins, Pacula, Paddock, and Chiesa (2004) argued that small and modest effect sizes found in universal programs, such as school-based drug and alcohol misuse prevention programs, can translate into practically important

economic benefits. Crutzen (2010) demonstrated that multicomponent initiatives can lead to smaller effects dissipated across inputs but also could lead to larger effects if synergies among inputs occurred. Crutzen proposed that an effect size needs to be considered in the context of five dimensions, namely, reach, efficacy, adoption, implementation, and maintenance. From the above perspectives, the public health impact of an intervention (such as KidsMatter, 2010, which is now being rolled-out to over 2000 schools across Australia) that has a small effect, but reaches a large group of people over an extended period of time, can still be high.

Limitations

In this study, we gathered questionnaire data from different categories of informants at four time points. The review of the strengths and limitations of types of data collection methods by Muijs (2006) highlighted that no single method of data collection is without limitations. Questionnaires, by design, take a broad perspective and may lack contextual sensitivity. Furthermore, self-report questionnaires may be coloured by socially desirable responses or self-reflective blind spots. The questionnaires used in our study were piloted for construct validity, all items retained in the final version of the questionnaire met acceptable criteria for reliability, and conservative statistical approaches that made limited assumptions about the distribution of the data were used. However, additional data collection methods, such as repeated observations of school-based actions as suggested by Muijs, are indicated to accompany the use of tools such as an Implementation Index.

A second limitation relates to the categorisation of schools into low- and high-implementation groups. As Durlak and DuPre (2008) pointed out, such categorisation within one study only identifies relative high and low implementation within that study. However, some degree of comparative, external perspectives across broadly diverse locations and types of schools (albeit within the same study) was provided by one of our participant cohorts, namely, the KidsMatter project officers. Importantly, as described by Coe (2002), replication of such studies can build an evidence base founded upon repeated patterns of items of interest.

Conclusions

In conclusion, we have argued that settings-based approaches to mental health promotion face many challenges associated with the diversity of participants' and sites' backgrounds, needs, and resources. Health promotion initiatives demand evaluation to provide evidence for ongoing commitment of resources. Evaluations can be of different types, including outcomes, resources, and process evaluations. This paper has focussed upon the latter type, using the Domitrovich et al. (2008) dimensions of fidelity, dosage, and delivery to inform the design of an evaluation tool, namely, the Implementation Index. The Index has proved capable of discriminating between schools' performance on a range of indicators of successful implementation. Furthermore, the Index has enabled the demonstration of a relationship between average- and high-implementing schools and improved student social and emotional competencies.

The findings reported in this paper directly respond to the concerns raised by Domitrovich and Greenberg (2000) about the lack of studies reporting relationships between processes and outcomes. The collection of both process information

with LCA and outcome information with HLM analysis of change over time provides an alternative method to a randomised control group experimental design, thus responding to the concerns raised by Weare and Nind (2011) about the quality of evaluation data in Australian studies of mental health promotion initiatives. Our study demonstrates an example of what Kam et al. (2003) proposed, where the evaluation of processes accompanies evaluation of outcomes, thus seeking complementarity, rather than dichotomising, the two evaluation types. This information can inform policies and plans for future interventions and evaluations.

Notes

1. For ease of reference, the term "parents" is used in this paper to refer to parents and other primary caregivers of children.
2. KidsMatter was undertaken in 101 schools, but one hospital-based school did not participate in the evaluation due to the challenges of collecting longitudinal questionnaire-based data from the school's transient students.
3. More detail about the aims and processes of KidsMatter can be obtained from the KidsMatter website, www.kidsmatter.edu.au
4. Full details of item selection, verification, and factor analysis can be found in Dix et al. (2010).
5. A few schools contained less than 76 students.
6. The display in Figure 6 suggests that students in high-implementing schools started with higher SEC than students in low-implementing schools. However, an extension to the HLM rejected INDEX as having a significant influence on the intercept ($p > .05$).

Notes on contributors

Helen Askell-Williams (PhD) is the Director of the Flinders University Educational Futures Research Institute. She has worked on collaborative research projects that have investigated teachers' and learners' knowledge and wellbeing. Her areas of research interest include students' motivational, cognitive, and metacognitive strategy knowledge and mental health interventions in educational settings. Helen lectures in the fields of educational psychology and building positive mental health.

Katherine L. Dix (PhD) is a Lecturer in the School of Education at Flinders University working with postgraduate students and undertaking research. Her research in the area of mental health promotion is characterised by a holistic approach in the design, analysis, and reporting of large-scale longitudinal projects conducted in primary school and early childhood settings. In addition, she has comprehensive experience in the use of specialised statistical packages and lectures in the field of Educational Research Methods.

Michael J. Lawson (PhD) is a Professor in the School of Education at Flinders University and is a member of the Flinders University Educational Futures Research Institute. His research has focused on the use of strategies for learning in classroom and study situations. In recent years, this research has been concerned with applications in mathematical problem solving, vocabulary acquisition in language learning, and the facilitation by teachers of students' use of learning strategies. In addition, Mike has been involved as a member of teams that have evaluated the effects of interventions on mental health for students in primary schools and early childhood centres.

Phillip T. Slee (PhD) is a Professor in Human Development in the School of Education at Flinders University and is a member of the Flinders University Educational Futures Research Institute. He is co-director of the Flinders Centre for Student Wellbeing and Prevention of Violence. Phillip is a trained teacher and registered psychologist. He has worked on a number of research projects concerned with school bullying, school violence prevention, and mental health promotion.

References

Askell-Williams, H., Lawson, M.J., & Slee, P.T. (2009). Venturing into schools: Locating mental health initiatives in complex environments. *The International Journal of Emotional Education*, *1*(2), 14–33.

Barry, M.M., & Jenkins, R. (2007). *Implementing mental health promotion*. Edinburgh, UK: Churchill Livingstone.

Bellg, A.J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D.S., Ory, M., . . . Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*, *23*, 443–451. doi:10.1037/0278-6133.23.5.443

Biglan, A., Mrazek, P.J., Carnine, D., & Flay, B.R. (2003). The integration of research and practice in the prevention of youth problem behaviors. *American Psychologist*, *58*, 433–440. doi:10.1037/0003-066X.58.6-7.433

Caulkins, J.P., Pacula, R.L., Paddock, S., & Chiesa, J. (2004). What we can-and cannot-expect from school-based drug prevention. *Drug and Alcohol Review*, *23*, 79–87. doi:10.1080/09595230410001645574

Clift, S., & Jensen, B.B. (Eds.). (2005). *The health promoting school: international advances in theory, evaluation and practice*. Copenhagen, Denmark: The Danish University of Education Press.

Coe, R. (2002, September). *It's the effect size, stupid: What effect size is and why it is important*. Paper presented at the Annual Conference of the British Educational Research Association, Exeter, UK.

Collaborative for Academic Social and Emotional Learning. (2008). *Connecting social and emotional learning with mental health*. Retrieved from http://casel.org/wp-content/uploads/2011/04/SELandMH.pdf

Collaborative for Academic Social and Emotional Learning. (2011a). *Programs in action*. Retrieved from http://casel.org/in-schools/selecting-programs/programs-in-action/

Collaborative for Academic Social and Emotional Learning. (2011b). *What is SEL? Skills & Competencies*. Retrieved from http://casel.org/why-it-matters/what-is-sel/

Colquhoun, D. (2005). Complexity and the health promoting school. In S. Clift & B.B. Jensen (Eds.), *The health promoting school: International advances in theory, evaluation and practice* (pp. 41–53). Copenhagen, Denmark: The Danish University of Education Press.

Conduct Problems Prevention Research Group. (2010). The effects of a multiyear universal social-emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology 78*, 156–168. doi:10.1037/a0018607

Crutzen, R. (2010). Adding effect sizes to a systematic review on interventions for promoting physical activity among European teenagers. *International Journal of Behavior, Nutrition and Physical Activity 7*, 29. doi:1186/1479-5868-7-29

Dane, A.V., & Schneider, B.H. (1998). Program integrity in primary and early secondary prevention: Are implementation efforts out of control? *Clinical Psychology Review*, *18*, 23–45. doi:10.1016/S0272-7358(97)00043-3

Department for Children, Schools and Families. (2009). *Promoting and supporting positive behaviour in primary schools. Developing social and emotional aspects of learning (SEAL)*. Nottingham, UK: Author.

Devaney, E., O'Brien, M.U., Resnik, H., & Weissberg, R.P. (2006). *Sustainable schoolwide social and emotional learning (SEL)*. Chicago, IL: University of Illinois.

Dix, K.L., Keeves, J.P., Slee, P.T., Lawson, M.J., Russell, A., Askell-Williams, H., . . . Spears, B. (2010). *KidsMatter primary evaluation: Technical report and user guide*. Adelaide, Australia: Shannon Research Press.

Domitrovich, C.E., Bradshaw, C.P., Poduska, J.M., Hoagwood, K., Buckley, J.A., Olin, S., . . . Lalongo, N.S. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools. A conceptual framework. *Advances in School Mental Health Promotion*, 1(3), 6–28.

Domitrovich, C.E., & Greenberg, M.T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation*, 11, 193–221. doi:10.1207/S1532768XJEPC1102_04

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. doi:310.1007/s10464-10008-19165-10460

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. doi:10.1111/j.1467-8624.2010.01564.x

Elias, M.J., Zins, J.E., Graczyk, P.A., & Weissberg, R.P. (2003). Implementation, sustainability, and scaling up of social-emotional and academic innovations in public schools. *School Psychology Review*, 32, 303–319.

European Centre for Educational Resilience & Socio-Emotional Health. (2009). Retrieved from http://www.um.edu.mt/edres

European Network for Social and Emotional Competence. (2009). Retrieved from http://www.enseceurope.org/

Foxcroft, D.R., & Tsertsvadze, A. (2011). Universal multi-component prevention programs for alcohol misuse in young people (Review). *Cochrane Database of Systematic Reviews*, 9:CD009307, 1–69.

Fullan, M. (2007). *New meaning of educational change* (4th ed.). New York, NY: Columbia University.

Garson, G.D. (2009). Structural equation modeling. *Statnotes*: *Topics in Multivariate Analysis*. Retrieved from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm

Giesen, F., Searle, A., & Sawyer, M. (2007). Identifying and implementing prevention programmes for childhood mental health problems. *Journal of Paediatrics & Child Health*, 43, 785–789. doi:10.1111/j.1440-1754.2007.01196.x

Graetz, B., Littlefield, L., Trinder, M., Dobia, B., Souter, M., Champion, C., . . . Cummins, R. (2008). KidsMatter: A population health model to support student mental health and well-being in primary schools. *International Journal of Mental Health Promotion*, 10(4), 13–20.

Greenberg, M.T. (2011, June-July). *Promoting wellbeing in schools: Current status and future challenges*. Keynote address at the bi-annual conference of the European Network for Social and Emotional Competence in Children, Manchester, UK.

Grissom, R.J., & Kim, J.J. (2012). *Effect sizes for research* (2nd ed.). New York, NY: Routledge.

Hallam, S. (2009). An evaluation of the Social and Emotional Aspects of Learning (SEAL) programme: Promoting positive behaviour, effective learning and wellbeing in primary school children. *Oxford Review of Education*, 35, 313–330. doi:10.1080/03054980902934597

Hox, J.J., & Bechger, T.M. (1998). An introduction to structural equation modeling. *Family Science Review*, 11, 354–373.

Humphrey, N., Lendrum, A., & Wigelsworth, M. (2010). *Social and emotional aspects of learning (SEAL) programme in secondary schools: National evaluation* (No. RB049). Retrieved from https://www.education.gov.uk/publications//eOrderingDownload/DFE-RB049.pdf

Hutchinson, A.D., & Wilson, C. (2011). Improving nutrition and physical activity in the workplace: A meta-analysis of intervention studies. *Health Promotion International*. *Advance online publication*. doi:10.1093/heapro/dar035

Jaycox, L.H., McCaffrey, D.F., Ocampo, B.W., Shelley, G.A., Blake, S.M., Peterson, D.J., . . . Kub, J.E. (2006). Challenges in the evaluation and implementation of school-based prevention and intervention programs on sensitive topics. *American Journal of Evaluation*, *27*, 320–336. doi:10.1177/1098214006291010

Kam, C.-M., Greenberg, M.T., & Walls, C.T. (2003). Examining the role of implementation quality in school-based prevention: Using the PATHS curriculum. *Prevention Science*, *4*, 55–63. doi:10.1023/A:1021786811186

KidsMatter. (2009). *KidsMatter primary framework*. Retrieved from http://www.apapdc.e-du.au/kidsmatter/

KidsMatter. (2010). *KidsMatter in 2011 and beyond*. Canberra, Australia: Commonwealth of Australia. Retrieved from http://www.kidsmatter.edu.au/kidsmatter-overview/kidsmatter-in-2011

Kids Matter Early Childhood. (n.d.). *Welcome to KidsMatter Early Childhood*. Retrieved from http://www.kidsmatter.edu.au/ec/

Lee, C., August, G.J., Realmuto, G.M., Horowitz, J.L., Bloomquist, M.L., & Klimes-Dougan, B. (2008). Fidelity at a distance: Assessing implementation fidelity of the early–risers prevention program in going-to-scale intervention trial. *Prevention Science*, *9*, 215–229. doi:10.1007/s11121-008-0097-6

Melde, C., Esbensen, F.-A., & Tusinski, K. (2006). Addressing program fidelity using onsite observations and program provider descriptions of program delivery. *Evaluation Review*, *30*, 714–740. doi:10.1177/0193841X06293412

Melhuish, E., Belsky, J., Anning, A., Ball, M., Barnes, J., Romaniuk, H., . . . The NESS Research Team. (2007). Variation in community intervention programmes and consequences for children and families: The example of Sure Start Local Programmes. *The Journal of Child Psychology and Psychiatry*, *48*, 543–551. doi:10.1111/j.1469-7610.2007.01705.x

MindMatters. (n.d.). *About MindMatters*. Retrieved from http://www.mindmatters.edu.au/about/about_landing.html

Mitchell, J., Palmer, S., Booth, M., & Davies, G.P. (2000). A randomised trial of an intervention to develop health promoting schools in Australia: The South Western Sydney study. *Australian and New Zealand Journal of Public Health*, *24*, 242–245. doi:10.1111/j.1467-842X.2000.tb01563.x

Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation*, *12*, 53–74. doi:10.1080/13803610500392236

Mukoma, W., & Flisher, A.J. (2004). Evaluations of health promoting schools: A review of nine studies. *Health Promotion International*, *19*, 357–368. doi:10.1093/heapro/dah309

Muthén, L.K., & Muthén, B.O. (2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Authors.

Oshima, I., Cho, N., & Takahashi, K. (2004). Effective components of a nationwide case management program in Japan for individuals with severe mental illness. *Community Mental Health Journal*, *40*, 525–537. doi:10.1007/s10597-004-6126-4

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London, UK: Sage.

Payne, A.A. (2009). Do predictors of the implementation quality of school-based prevention programs differ by program? *Prevention Science*, *10*, 151–167. doi:10.1007/s11121-008-0117-6

Raudenbush, S.W., Bryk, A.S., Cheong, Y.F., & Congdon, R.T. (2004). *HLM6. Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Resnick, L.B. (2010). Nested learning systems for the thinking curriculum. *Educational Researcher*, *39*, 183. doi:10.3102/0013189X10364671

Rootman, I. (2001). Introduction to the book. In I. Rootman, M. Goodstadt, B. Hindman, D.V. McQueen, L. Potvin, J. Springett, & E. Ziglio (Eds.), *Evaluation in health promotion*. *Principles and perspectives* (pp. 3–6). Copenhagen, Denmark: World Health Organization Regional Publications, European Series 92.

Rosenthal, R., & DiMatteo M.R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Reviews of Psychology*, *52*, 59–82. doi:10.1146/annurev.psych.52.1.59

Shek, D.T.L., Sun, R.C.F., & Kan, V.W.M. (2009). Full implementation of the Secondary 1 Program of Project P.A.T.H.S.: Observations based on the co-walker scheme. *The Scientific World Journal*, *9*, 982–991.

Slee, P.T., Lawson, M.J., Russell, A., Askell-Williams, H., Dix, K.L., Owens, L., . . . Spears, B. (2009). *KidsMatter evaluation final report*. Adelaide, Australia: Flinders University.

Van Der Heyden, A.M., Witt, J.C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology*, *45*, 225–256. doi:10.1016/j.jsp.2006.11.004

Waks, L.J. (2007). The concept of fundamental educational change. *Educational Theory*, *57*, 277–295. doi:10.1111/j.1741-5446.2007.00257.x

Weare, K. (2010). Mental health and social and emotional learning: Evidence, principles, tensions, balances. *Advances in School Mental Health Promotion*, *3*, 5–17.

Weare, K., & Nind, M. (2011). Mental health promotion and problem prevention in schools: What does the evidence say? *Health Promotion International*, *26* (SUPPL. 1), i29–i69. doi:10.1093/heapro/dar075

World Health Organization. (2007). *What is a health promoting school?* Retrieved from http://www.who.int/school_youth_health/gshi/hps/en/index.html

World Health Organization and Health and Welfare Canada. (1986). *Ottawa Charter for Health Promotion*. Ottawa, Canada: World Health Organization.

World Health Organization Europe. (2006, June). *Opening address*. Paper presented at the fifth workshop on practice of evaluation of the Health Promoting School – concepts, indicators and evidence, Sigriswil, Switzerland.