The impact of trial stage, developer involvement and international transferability on universal social and emotional learning programme outcomes: A meta-analysis

Wigelsworth, M., Lendrum, A., Oldfield, J., Scott, A., ten Bokkel, I., Tate, K., & Emery, C.

The impact of trial stage, developer involvement and international transferability on universal social and emotional learning programme outcomes: A meta-analysis

Abstract

This study expands upon the extant prior meta-analytic literature by exploring previously theorised reasons for the failure of school-based, universal social and emotional learning (SEL) programmes to produce expected results. Eighty-nine studies reporting the effects of school-based, universal SEL programmes were examined for differential effects on the basis of: 1) stage of evaluation (efficacy or effectiveness); 2) involvement from the programme developer in the evaluation (led, involved, independent); and 3) whether the programme was implemented in its country of origin (home or away). A range of outcomes were assessed including: social-emotional competence, attitudes towards self, pro-social behaviour, conduct problems, emotional distress, academic achievement and emotional competence. Differential gains across all three factors were shown, although not always in the direction hypothesised. The findings from the current study demonstrate a revised and more complex relationship between identified factors and dictate major new directions for the field.

Key words: meta-analysis, socio-emotional, efficacy, developer, transferability

Literature Review

There is an emerging consensus that the role of the school should include supporting children's emotional education and development (Greenberg, 2010; Weare, 2010). This is often accomplished through the implementation of universal social and emotional learning (SEL) programmes which aim to improve learning, promote emotional well-being, and prevent problem behaviours through the development of social and emotional competencies (Elias et al., 2001; Greenberg et al., 2003).

What is SEL?

Social and emotional learning (SEL) is represented by the promotion of five core competencies: self-awareness; self-management; social awareness; relationship skills; and responsible decision-making (Collaborative for Academic, Social and Emotional Learning, 2002). Although a broad definition serves to encompass many aspects for the effective promotion of SEL, it does little to differentiate or identify 'essential ingredients' of a programme of change. As a result, SEL is implemented through a variety of formats, differing levels of training and support, varying degrees of intensity, and variation in regards to the relative importance placed in each of the five core competencies. However, most SEL programmes typically feature an explicit taught curriculum, delivered by teachers (with or without additional coaching and technical support), and are delivered during school hours (examples of SEL programmes can be seen at casel.org).

Currently, a wide range of SEL programmes feature in schools and classrooms across the world, including in the USA (e.g. Greenberg, Kusche, Cook, & Quamma, 1995), Australia (e.g. Graetz et al., 2008), across Europe (e.g. Holsen, Smith, & Frey, 2008), and in the UK (e.g. DfES, 2007). Although a poor tool for assessing the complexity of any specific curriculum intervention or context, meta-analytic approaches are required to support the hitherto theoretical assumption of the universality of teachable SEL competencies. Indeed, recent meta-analyses in the United States (Durlak, Weissberg, Dymicki, Taylor, & Schellinger, 2011) and the Netherlands (Sklad, Diekstra, Ritter, & Ben,

2012) have been used to suggest that high quality, well-implemented universal SEL interventions, designed to broadly facilitate a range of intra- and inter-personal competencies, can lead to a range of salient outcomes, including improved social and emotional skills, school attitudes and academic performance, and reduced mental health difficulties (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Sklad, Diekstra, Ritter & Ben, 2012; Wilson & Lipsey, 2007). However, individual SEL programmes are not always able to produce the same impressive results indicated by these meta-analyses when adopted and implemented by practitioners in schools (Social and Character Development Research Consortium, 2010).

Research in prevention science suggests a number of possible reasons for this discrepancy, including implementation failure (Durlak & DuPre, 2008), a reliance on the results of 'early' trials focusing on the internal logic of intervention, rather than their 'real world' applicability (Flay, et al., 2005), developer involvement in trials (Eisner, 2009) and a lack of cultural transferability of interventions (Castro, Barrera, & Martinez, 2004). Although implementation fidelity is now recognised as an important feature in the successful delivery of SEL programmes (included in Durlak et al.'s 2011 meta-analysis), there has been no similar empirical consideration of the other factors. Underlying such explanations is an implicit assumption of a degree of invariance or 'treatment' approach in the implementation of SEL programmes. Many consumers of educational research will recognise a 'medical model' of evaluation (typically involving experimental designs), an approach which is not without debate (for a brief summary see Evans and Benefield (2001)). Accordingly, prior research in educational evaluation has noted that such an approach is potentially limited, as associated methodologies for investigation (neatly described by Elliott and Kushner (2007) as the, "statistical aggregation of gross yield" (p.324)) will fail to capture complexities of the interactions within specific contexts required to explain findings. Indeed, lack of process and implementation has been noted in this particular field (Lendrum & Humphrey, 2010). However, suggested alternate directions (e.g. anthropological, illuminative, case study (ibid)) can fail to capture the prevalence or magnitude of trends, and, in their own way, also fail to uncover important lessons as to the successful

implementation of educational interventions. Therefore, there is an opportunity to extend prior influential work (i.e. Durlak, Weissberg, Dymicki, Taylor, & Schellinger, 2011; Sklad, Diekstra, Ritter, & Ben, 2012) utilising meta-analytic approaches, to examine key indicators potentially influencing successful outcomes, and in doing so, to consider extent to which such techniques are useful in this context.

Key indicators

In attempting to explain the high degree of variation between programmes in achieving successful outcomes for children, this article now considers the rationale for exploring theoretically important (e.g. as conceptualised by Lendrum and Wigelsworth (2013)), but often ignored factors, and hypothesises their likely effect on SEL programme outcomes.

Stage of evaluation: Efficacy vs. effectiveness

An important (and often promoted) indicator as to the potential success of a programme is its history of development and evaluation. Ideally, an intervention should be tested at several stages between its initial development and its broad dissemination into routine practice (Greenberg, Domitrovich, Graczyk, & Zins, 2005) and frameworks have been provided in recent literature to enable this. For instance, drawn from evaluations of complex health interventions, Campbell et al. (2000) provide guidance on specific sequential phases for developing interventions: developing theory (pre-phase), modelling empirical relationships consistent with intended outcomes (phase I), exploratory trialling (phase II), randomised control trials under optimum conditions (phase III), and long term implementation in uncontrolled settings (phase IV). An intervention should pass through all phases to be considered truly effective and evidence-based (Campbell et al., 2000).

An important distinction in Campbell et al.'s framework is the recognition that interventions are typically first 'formally' evaluated under optimal conditions of delivery (phase III) (more broadly referred to as efficacy trials (Flay, 1986)), such as with the provision of highly-trained and carefully

supervised implementation staff.   Subsequent to this, a programme may be tested under more 'real world' or naturalistic settings, using just the staff and resources that would be normally available. This is aligned to Campbell et al.'s phase IV and is commonly known as an effectiveness trial (Dane & Schneider, 1998; Flay et al., 2005).   Although both types of trial may utilise similar overarching research designs (e.g. quasi-experimental or randomised designs), they do seek to answer different questions about an intervention.   Whereas efficacy studies are typically conducted to demonstrate the efficacy and internal validity of a programme, effectiveness is used to test whether and how an intervention works in real-world contexts (Durlak, 1998; Greenberg et al., 2005).   This also allows identification of factors that may influence the successful adoption, implementation, and sustainability of interventions when they 'go to scale' (Greenberg, 2010), for instance by highlighting additional training needs or workload allocation pressures.   Thus, a programme that demonstrates success at the efficacy stage may not yield similar results under real world conditions.   Indeed, research indicates that practitioners in 'real-world' settings are generally unable to duplicate the favourable conditions and access the technical expertise and resources that were available to researchers and programme developers at the efficacy stage (Greenberg et al., 2005; Hallfors & Godette, 2002) and thus fail to implement programmes to the same standard and achieve the same outcomes (Durlak & DuPre, 2008).

An example of this distinction is an effectiveness trial of the Promoting Alternative Thinking Strategies (PATHS) curriculum in the Netherlands (Goossens et al., 2012) (a programme closely aligned to the general descriptors provided in the introduction).   The study failed to replicate expected effects demonstrated in earlier efficacy trials (e.g. Greenberg et al., 1995).   Demonstrating an implementation strategy that allowed for high degrees of adaptation, the authors of the study concluded that the implementation strategy adopted was "not a recipe for effective prevention of problem behavior on a large scale" (p245).

Despite calls for prevention programmes to be tested in multiple contexts before they are described as 'evidence-based' (Kumpfer, Magalhães, & Xie, 2012), there is, as yet little clarification in the SEL literature (including major reviews) regarding the stage of evaluation of programmes and whether those classified as 'successful' or 'exemplary', have achieved this status on the basis of efficacy alone or have also undergone effectiveness trials.

The involvement of programme developers

There are many logical reasons why the developer of a specific SEL intervention would also conduct evaluation trials, especially during efficacy phase (as above). However, there is evidence from associated fields to suggest that the involvement of the programme developer with an evaluation may be associated with considerably greater outcomes (Eisner, 2009). For example, in a review of psychiatric interventions, studies where a conflict of interest was disclosed (e.g. programme developers were directly involved in the study) were nearly five times more likely to report positive results when compared to truly 'independent' trials (Perlis et al., 2005). Similarly, an independent effectiveness study of Project ALERT, a substance abuse prevention programme, failed to find positive outcomes, despite successful efficacy and effectiveness studies conducted by the programme developer (St. Pierre et al., 2005).

Eisner (2009) posits two possible explanations for this phenomenon. The *cynical view* proposes that the more favourable results in developer-led trials stem from systematic biases that influence decision-making during a study. Alternatively, the *high fidelity view* argues that implementation of a given intervention is of a higher quality in studies in which the programme developer is involved, leading to better results. In either case, developer involvement leads to an inflation of outcome effect sizes compared to those which might be expected from 'real world' implementation of a given programme. The obvious consequence of such an effect is the inherent difficulty in replication of

expected effects in any wider dissemination or 'roll out' of the programme. If the intended outcomes of an intervention may only be achieved if the programme developer is available to enforce the highest levels of fidelity, then its broad dissemination and sustainability across multiple setting is unlikely to be feasible. Despite Eisner's observations, recent reviews and meta-analyses of SEL programmes do not distinguish between evaluations conducted by external researchers and those led by, or with the involvement of, programme developers or their representatives.

Cultural transferability

Issues of cultural transferability have particular implications for SEL programmes. This is because perceived success in the context of the USA (around 90% of the studies included in Wilson and Lipsey's (2007) and Durlak et al.'s (2011) reviews originated there) has resulted in rapid global dissemination and adoption of SEL programmes. For instance, PATHS (Greenberg & Kusché, 2002), Second Step (Committee for Children, 2011), and Incredible Years (Webster-Stratton, 2011) have been adopted and implemented across the world (e.g. Henningham, 2013; Holsen et al., 2008; Malti, Ribeaud & Eisner, 2011).

International transfers of programmes provide valuable opportunities to examine aspects of implementation, with special regard to the fidelity-adaptation debate (Ferrer-Wreder, Adamson, Kumpfer, & Eichas, 2012). This is because a major factor in the successful transportability of interventions is their adaptability (Castro, Barrera, & Martinez, 2004). Accepting the view that successful outcomes may rely on at least some adaptation to fit with cultural needs, values, and expectations of the adopters *within* countries of origin (Castro, Barrera & Martinez, 2004), the complexities of international transferability *across* countries becomes apparent. Adaptations vary, and although surface level changes (e.g. modified vocabulary, photographs, or names) may be beneficial and enhance cultural acceptability, deeper structural modifications (e.g. different pedagogical approaches or modified programme delivery), and may compromise the successful implementation of the critical components of an intervention. This may have serious negative

consequences, to the extent that change is not triggered and the outcomes of the programme are not achieved. Indeed, there is arguably the potential for programmes to be adapted to cultural contexts to such an extent that they become, in effect, new programmes requiring re-validation, ideally through the use of an evidence framework such as Campbell et al.'s (2000) in order to test the underlying programme theory and internal validity.

Unsurprisingly, findings for adopted or imported programmes are mixed. For instance, a number of studies report null results for 'successful' USA programmes transported into the UK including anti-bullying programmes (Ttofi, Farrington, & Baldry, 2008), sex, and drugs education (Wiggins et al., 2009). Some programmes show mixed success when transferred internationally (such as PATHS), reporting varying levels of success from null results in England (Little et al., 2012) to mixed effects in Switzerland (treatment effects were identified for only some outcomes) (Malti, Ribeaud, & Eisner, 2011). Conversely, the SEL programme 'Second Step' (Committee for Children, 2011) has been shown to have positive effects across several sites in the USA (e.g. Cooke et al., 2007; Frey, Nolen,Edstrom & Hirschstein, 2005) and in Europe (Holsen et al., 2008). Therefore, there are questions as to the extent to which programmes can achieve the same intended outcomes when transported to countries with different education systems, pedagogical approaches, and cultural beliefs.

Research Questions and Aims

This study is the first of its type to examine the potential effects of the identified factors on the outcomes of universal school-based programmes. To date, previous reviews have been limited; reporting on a limited palate of intervention type (Durlak et al., 2011), main effects of SEL programmes only (Sklad et al., 2012), or have examined the effect of programme variables themselves (Durlak et al., 2011). The purpose of the current study is to build upon this prior work to assess to what extent meta-analytical techniques can help explain inconsistencies in demonstrating

positive programme outcomes. Given the variability identified in the proceeding review, the relative usefulness of meta-analytical techniques will also be considered.

As previous reviews have already established positive main effects across a variety of skills and behaviours, our hypotheses focus on the differential effects of the categories identified through the literature review, specifically;

1) Studies coded as 'efficacy' will show larger effect sizes compared to those coded as 'effectiveness'

2) Studies in which the developer has been identified as leading or being involved will show larger effect sizes in relation to independent studies

3) Studies implemented within the country of development (home) will show larger effect sizes than those adopted and implemented outside country of origin (away)

Methods

A meta-analytic methodology was adopted to address the study hypotheses, in order to ensure an unbiased, representative and high quality process of review. 'Cochrane protocols for systematic reviews of interventions' (Higgins & Green, 2008) were adopted for the literature searches, coding, and analytical strategy. To address the common issue of comparing clinically diverse studies ('apples with oranges'), outcome categories were classified on the basis of prior work in the field (e.g. Durlak et al., 2011; Sklad et al., 2012; Weare & Nind, 2010; Wilson & Lipsey, 2007) and analysed seperately.

For the purposes of the current study, we have co-opted Denham's (2005) framework of social and emotional competence. This is an extremely close fit to the five SEL competencies and provides an additional layer of specification to ensure specific outcomes are identifiable alongside broader measures of SEL.

***Table 1 ***

Literature search

Four search strategies were used to obtain literature. First, relevant studies were identified through searches of major scientific databases, specifically; ASSIA, CINALAH, Cochrane database of systematic reviews, EMBASE, ERIC, MEDLINE, NICE, PsychINFO, and additional web searching using Google Scholar. Second, a number of journals most likely to contain SEL based publications were also searched. For instance; *Prevention Science*, *Psychology in the Schools*, and *School Psychology Review*. Third, organisational websites promoting SEL were searched to identify additional studies (e.g. casel.org). For all searches, the following key terms were used in different combinations to help maximise the search results:

"*SEL, social, emotional, wellbeing, mental health, intervention, programme, promotion, initiative, pupil, school, impact, effect, outcome, evaluation, effectiveness, scale, efficacy, pilot, independent, developer*".

Fourth, the reference list of each identified study was reviewed.

Inclusion criteria

Studies eligible for the meta-analysis were: a) written in English; b) appeared in published or unpublished form between 01 January 1995[1] and 01 January 2013; c) detailed an intervention that included the development of one or more core SEL components as defined by Denham (2005); d) delivered on school premises, during school hours; e) delivered to students aged 4 – 18 years; f) detailed an intervention that was universal (i.e. for all pupils, regardless of need); g) included a control group; h) reported sufficient information for effect sizes to be calculated for programme effect.

---

[1] 1995 denotes the release of Daniel Goleman's book "Emotional Intelligence", which marked the mainstream acceptance of SEL related competencies and associated interventions under the current rubric.

Exclusion criteria

Studies with the sole intended outcome of either academic attainment or physical health (e.g. drug use or teen pregnancy) were excluded. This criterion did not include studies where academic attainment was presented a distal outcome (e.g a secondary or tertiary measure after capturing social and emotional outcomes), reflecting that academic attainment is seen as a distal outcome derived from improvements in social and emotional skills (see Zins, Bloodworth, Weissberg, & Walberg, 2007). Studies that applied universal interventions to small groups within classes (e.g. targeted subgroup) were also excluded. Studies without appropriate descriptive statistics (i.e means and standard deviations) were also excluded.

A total of 24,274 initial records were identified. Inclusion / exclusion criteria were applied to these brief records, and as multiple researchers were used to search for literature,   large numbers of duplicates were also removed. After this process, 327 initial records remained, in which the full text was required to assess eligibility. From these full texts, a further 238 studies were excluded, leaving a final 89 studies to be retained for coding and analyses.

Coding

Methodological variables

To account for methodological variability affecting outcomes, three methodological variables were coded. This included:

Time of assessment: post-test (0-6 months, follow up (7-18 months), extended follow up (over 18 months).

 Period of schooling: early primary (4-6years), late primary (7-11years), early secondary (12-14years), and late secondary onwards (15-18years).

Type of design: randomised, cluster randomised, wait-list (non randomised), matched comparison, other.

Two dichotomous variables assessing implementation were also coded; implementation integrity (implementation reported: yes / no) and, implementation issues (issues with implementation: yes / no).

Independent variables

In order to address the current study's hypotheses, studies were coded by their stage of evaluation (efficacy or effectiveness), any involvement of the intervention author (developer involvement), and by place of origin and implementation (cross-cultural transferability):

Efficacy/ Effectiveness: Studies were coded dichotomously (efficacy or effectiveness) according to whether trial schools or teachers had received, support, training, staff or resources that would not normally be available to the organisation purchasing/ adopting the intervention (efficacy) (Flay, 1986; Dane & Schneider 1998), or whether the intervention had been conducted in naturalistic settings without developer/ researcher support , using just the staff and resources that would be normally available (ibid).

Developer involvement: Studies were coded as 'led', 'involved', or 'independent' according to whether a study was identified as receiving developer support. This was accomplished through the list of authors, methods section, and any declaration of competing interests in the individual papers. Where there was no citation of a programme developer studies were crossed referenced across programme authors (where known). Where a programme a developer was not present in the author list, no developer support was referenced in the methods section (or, indeed, independence was declared), and there was a declaration of competing interest, the programme was coded as 'independent'. This is on the basis that were a programme author involved in cases coded as 'independent', their contribution was minor enough to warrant omission from the reporting of the study.

Cross-cultural transferability: Studies were coded dichotomously (home or away) according to whether the trial had taken place in the country of programme development. We assessed transferability by national boundaries only (e.g. different States within the USA were considered

'home' for a USA programme). A programme that had received surface level adaptations (e.g. translation) was considered to be 'transported' programme and classified as 'away'. We did not find any evidence of a programme being so fundamentally altered (i.e. change to its internal logic model) as to classify as a new 'home' programme.

Outcome variables

Outcome variables were classified on the basis of prior work in this area (e.g. Durlak et al., 2011; Sklad et al., 2012; Weare & Nind, 2010; Wilson & Lipsey, 2007) with some minor adaptation to fit with the obtained literature and the adopted conceptual framework from Denham (2005). Seven different outcomes were identified:

1) Social-emotional competence (SEC): This category included measurement of general socio-emotional competency that incorporated both the emotional and relational aspect of Denham's model. Examples include total scores on omnibus measures (either self, peer or teacher rated) such as the 'Social Skills Improvement System' (Gresham & Elliot, 1990), and measures of interpersonal problem-solving/ conflict resolution behaviour.

2) Attitudes towards self (ATS): Outcomes relating exclusively to the skills of self-awareness were classified in this category. This included measures of self-esteem, self-concept, and general attitudes towards self, measured by instruments such as the 'Student Self Efficacy Scale' (Jinks & Morgan, 1999).

3) Pro-social behaviour (PSB) – This category was used to classify all measure of behaviours intended to aid others, identified as social awareness and/or social problem solving in Denham's model. Examples include self, peer, and teacher rated social behaviour and behavioural observations. An example measure includes the pro-social behaviour subscale of the 'Strengths and Difficulties Questionnaire' (Goodman, 2001).

In addition to Denham's framework, outcomes common to the logic model of the interventions were also coded;

4) Conduct problems (CP): Observational, self, peer, or teacher rated anti-social behaviour, externalised difficulties, bullying, aggression, or physical altercation was included in this category. An example instrument categorised as CP is the 'Child Behaviour Checklist' (Achenbach & Edelbrock, 1983).

5) Emotional distress (ED): Outcomes grouped in this category were representative of internalised mental health issues. Examples include self-rated, depression, stress and/or anxiety, measured by instruments such as the 'Beck Depression Inventory' (Beck, Steer, & Brown, 1996).

6) Academic Attainment (AA): Measures of attainment were judged through the use of reading tests, other standardised academic achievement measures (such as the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997)) or teacher-rated academic competence.

7) Emotional Competence Only (ECO): An additional category was coded to include measures exclusively associated with the internal domains related to emotional competency. Measures that assessed emotional skills (such as the Emotion Regulation Checklist (Shields & Cicchetti, 1997)) were included in this category, in order to differentiate some of the more generalised SEC inventories which provide single scores for both intra and inter personal skills.

Coding reliability: Coding of study characteristics (e.g. methodological characteristics) and outcomes data were completed by a small team of doctoral students who were familiar with the literature on social and emotional skills. All coders received training, which included practice double-coding of exemplar literature. Coders were allowed to proceed once they had reached full agreement with the lead authors. A training manual and online coding system were also provided to ensure consistency in coding. Coding consistency was monitored, and some variation in agreement in later coding led the lead authors to review the full dataset. The lead authors reached 100% agreement for the final coding selection.

Analytical strategy

Standard meta-analytic protocol as specified by Higgins and Green (2008) was followed for the analytical strategy and is detailed below. As data were derived from different sources, all scores for individual studies were standardised and calculated so that positive values indicated a favourable result for intervention groups compared to control (e.g. so measures of reduction in problem behaviours did not confound the analyses). All analyses were conducted using meta-analytic software ('MetaEasy' v1.0.2. Kontopantelis & Reeves, 2009).

For individual studies reporting more than one intervention, results were treated as multiple, separate entries. Results examining either: a) just subgroups (e.g. those identified with special educational needs); b) multiple interventions beyond only the identified SEL intervention (e.g. bespoke academic add-ons), or; c) Comparisons of different intervention treatments with each other (i.e. no control group) were not included in the analyses. When a study reported more than one outcome from the same category (e.g. conduct problems), scores were averaged to obtain a single figure.

Determining whether one category of study was significantly different (α <.05) from another (i.e. efficacy vs. effectiveness), was calculated using technique known as proportional overlap. As described by Cumming and Finch (2005), this approach does not require a specialised understanding of meta-analytic techniques, and instead assesses how closely related two confidence intervals, such as those presented in figures 1-3 (see 'results'). Using the same approach, individual variables (e.g. main programme effects) were assessed on the basis that the 95% confidence intervals did not cross zero.

A low number of individual studies for some of the categories for analysis prevented statistical significance testing being carried out (at least 5 studies for each category are required to be confident that any significant difference is accurate (Hedges & Pigott, 2001)), but this did not prevent the comparison of effect sizes.

For the purposes of the current study, all effects sizes are reported as Hedge's adjusted *g* (Hedges & Olkin, 1985), as described in Kontopantelis and Reeves (2009). Hedge's *g* is almost identical to the

often used Cohen's *d*, but provides a superior estimate of effect when sample sizes are small. Both *g* and *d* are interpreted in exactly the same way.   Estimation was based on a random effects model (Dersimonian & Laird, 1986). This technique was considered particularly suitable for the current study, as it assumes that different studies estimate different, but related, treatment effects (Higgins & Green, 2008).

The degree of diversity in programme effects were examined using summary values, specifically the Q and I^2 statistics.   Q is used to assess whether there was statistically significant variability (or 'heterogeneity'), whereas I^2 provides a measure of the *degree* of heterogeneity, measured by a percentage (0%-100%).   25, 50, 75 are considered low, medium, high degrees of heterogeneity respectively (Higgins, Thompson, Deeks, & Altman, 2003). For instance, if category is shown to have a significant Q value, and high I^2 (>75%), it can be said to be associated with a great deal of difference in the how successful individual programmes are in that category.

Results

Descriptive Characteristics

A final sample of 89 studies that met the aforementioned inclusion criteria was included in the analyses. Table 2 summarises the salient features of the included studies. These figures are broadly consistent with the characteristics of studies included in previous reviews (Durlak et al., 2011; Sklad et al., 2012). However, the majority of studies (69%) reported on implementation, a higher proportion compared to previous reviews.

Table 2

In relation to the specific criteria of the study hypotheses, most studies were considered to be reporting efficacy based trials (69%), with the majority including some element of developer

involvement, either as lead (38%) or involved (28%). Unsurprisingly, the majority of studies were of

'home' programmes, (80%), mostly originating from the USA. These figures are shown in table 3.

Table 3

Main programme effects

Table 4

The grand means for each of the outcome variables of interest were statistically significant (all 95%

CI's did not pass zero). The magnitude of the effect varied by outcome type, with the largest effect

for measures of social-emotional competence (0.53), and the smallest in attitudes towards self

(0.17).   However,   this   particular   variable   represents   only   a   small   number   of   studies   (n=9).

Heterogeneity   amongst   studies   was   high,   with I^2 varying   between   43%-97%,   confirming   that

although SEL programmes can be seen, on average, to deliver intended effects, there was a high

degree of variation amongst individual studies.

Stage of evaluation: Efficacy vs. Effectiveness

Table 5

We predicted that studies coded as efficacy would show greater effects compared to studies coded

as effectiveness. This hypothesis was supported for 6 of the 7 outcome variables (mean difference

(MD) between conditions = 0.13). However, comparisons of proportional overlap (Cumming & Finch,

2005) showed only 4 of these outcomes reached statistical significance (PSB, CP, ED, & AA). The

largest   differences   were   seen   between   outcomes   measuring   behaviour,   specifically   pro-social

behaviour (MD =0.19) and conduct problems (MD =0.19). The smallest differences were detected

for the outcomes variables of attitudes towards self (MD =0.06) and emotional competence only

(mean difference =0.05), though these also reflect the outcome categories with the smallest number of included studies. Only outcomes classified as 'social emotional competence' showed greater effects when in the effectiveness condition, contrary to the stated hypothesis. Effect size and confidence intervals can be seen in figure 1. Of note is the high degree of heterogeneity across all outcome variables, as evidenced by the values of both the Q and I^2 statistics, with most studies being categorised as either 'medium' or 'high' (Higgins, Thompson, Deeks, and Altman, 2003), indicating a diversity of effect across studies.

Figure 1

Developer involvement

Table 6

We hypothesised that studies in which the developer had been identified as taking a lead would show greater effects in relation to independent studies. Taking into account the small sample size for attitudes towards self (n=4), and the very high degree of heterogeneity noted by the Q and I^2 statistics, the hypothesis was supported by only 2 of the 7 outcomes ('attitudes towards self' and 'emotional distress'), though these were not seen to be statistically significant. The mean difference between developer led and independent for these outcomes was 0.2 and 0.02 respectively. The outcome variables for pro-social behaviour, academic achievement, and emotional competence only showed the greatest effects at 'involved', whereas social and emotional competence and conduct problems showed the highest mean effect at 'independent'. Effect size and confidence in intervals can be seen in figure 2.

To further investigate Eisner's (2009) high fidelity hypotheses (implementation of a given intervention is of a higher quality in studies in which the programme developer is involved, leading to better results), a cross-tabulated analysis (developer involvement vs. issues with implementation) was conducted for all studies which reported implementation (n=61). No significant association between developer involvement and issues with implementation was found (χ^2 (2, n=61) = .633, p =

.718, Cramer's V = .104).   This suggests that differences in effect between categories of developer involvement is not explained by better implementation.

Figure 2

Table 7

Cross-cultural transferability

Table 8

We hypothesised that studies implemented within the same country they were developed would show greater effects than those transported abroad. This hypothesis was supported in 4 of the 7 outcome variables: (SEC (MD = 0.5); ATS (MD = 0.11); PSB (MD = 0.19); & ED (MD = 0.09). All four were seen to be statistically significant. Only one study reporting 'attitudes towards self' qualified as 'away' and therefore fit statistics were not available. For the conduct problems, academic achievement and emotional competence, more favourable effects were seen for studies coded as 'away'. Both the Q and I^2 statistics show a very large degree of inconsistency between studies, as well as a very small n for some of outcome variables. This is a likely explanation for the large confidence intervals demonstrated in figure 3.

Figure 3

Discussion

The purpose of the current study was to empirically investigate previously hypothesised factors that may explain differential programme effects in universal, school-based SEL programmes, specifically: 1) stage of evaluation (efficacy or effectiveness); 2) involvement from the programme developer in the evaluation; and 3) whether the programme was implemented in its country of origin. Findings from the current study present a more complex picture than that hypothesised in previous literature. These findings necessitate new thinking about the way these (and other) factors are examined, and highlight important implications in the world-wide implementation of these interventions. Each hypothesis is discussed in turn, followed by a consideration of the limitations of the current study and consideration of future directions for research.

1) Studies coded as 'efficacy' will show larger effect sizes compared to those coded as 'effectiveness'

Greater programme effects under efficacy conditions (consistent with the first hypothesis) were shown for all but one outcome variable (social-emotional competence (SEC)), (though only 4 of the 7 outcomes were statistically significant). Results indicate a trend towards greater effects when additional support, training, staff or resources are provided. This is consistent with previous findings from Beelmann and Lösel (2006) who found an increased effect (d = 0.2), if a programme had been delivered by study authors or research staff. This finding has implications for the scaling up of programmes as this implies a large scale over-representation of expected effects in 'real world' settings, especially as 69% of studies were coded as efficacy trials.

These findings may be interpreted as indicating that the higher levels of fidelity produce higher outcomes, and has underpinned the argument that 100% fidelity is to be strived for and adaptations to be avoided in order to achieve outcomes (Elliott & Mihalic, 2004). Whilst not entirely unreasonable, this involves two assumptions that should be considered before dismissing the considerable literature that supports the utility of some types of adaptations. Firstly, it assumes that the salient characteristics of schools that later adopt an intervention mirror those of the school where an efficacy trial took place. Such an assumption rejects the inherent diversity of the school environment. Natural variation in contexts is to be expected (Forman et al., 2013), and adaptations to the programme or the way in which it is implemented may be necessary to achieve the same 'goodness-of-fit' as seen in an efficacy trial. Research consistently demonstrates that such adaptations are to be expected when school-based interventions are adopted and implemented more broadly (Ringwalt et al., 2003). Secondly, there is the assumption that at efficacy stage an intervention is implemented with 100% fidelity. As a primary aim of an efficacy trial is to demonstrate the internal validity of an intervention, and that the context of implementation is optimised to maximise the achievement of outcomes, then it is possible that either, or both, the context and the programme are adjusted, however slightly, to support the demonstration of impact.

Such considerations do not account for the contrary result for the SEC outcome, which shows larger effects under the effectiveness condition. One promising explanation for this conflicting finding is offered by our current understanding of self-efficacy in programme delivery. Self-efficacy is underpinned by knowledge, understanding and perceived competence, which has been shown as a factor in promoting achievement of outcomes (Durlak & DuPre, 2008; Greenberg et al., 2005). Indeed, there is evidence to suggest that, for some interventions, greater effects are achieved when the programme is delivered by external facilitators, when compared to teachers (Stallard, 2014). Therefore, it is possible that an effectiveness trial can outperform efficacy conditions, only when there is a high degree of implementer confidence and/or skill. If this is the case, then there may be a differential ease by which universal, school-based SEL interventions are seen as achievable by the implementers (i.e. school staff). Programmes featuring promotions of general socio-emotional competency (that incorporate both the emotional and relational aspect of Denham's model), may be viewed as the most acceptable and therefore the programmes that are hypothesised to most benefit from inevitable adaptation. In conjunction with previous paragraph, this might imply that adaptation is preferable to fidelity, only when there is sufficient confidence and understanding of the intervention. As literature indicates that multiple factors may be a factor in explaining a reduction in effects, (Biggs, Vernberg, Twemlow, Fonagy, & Dill, 2008; Stauffer, Heath, Coyne, & Ferrin, 2012), this is a clear steer towards a closer consideration of differing aspects when evaluating programme implementation, requiring a broader application of research methodologies to investigate.

2) Studies in which the developer has been identified as leading or being involved will show larger effect sizes in relation to independent studies

The study hypothesis was supported by 2 of the 7 outcomes ('attitudes towards self' and 'emotional distress'), though these were not seen to be statistically significant. Dependent on the outcome variable measured, effects favour either involvement (pro-social behaviour, academic achievement, emotional competence) or independence (social-emotional competence, conduct problems).

Therefore, consideration of developer involvement alone is not sufficient to explain variation in programme outcomes. This result is at odds with findings from allied disciplines such as Psychiatry (Perlis et al. (2005), and Criminology (Petrosino & Soydan, 2005), which demonstrate clear difference in effect when considering developer involvement. For instance, Perlis et al. (2005) found that studies that declare a conflict of interest were 4.9 times more likely to report positive results. However, attributing effects directly to the involvement of programme developers is questionable (Eisner, 2009). This is because developer involvement is an indirect indicator of other aspects of study design (e.g. experimenter expectancy effects (Luborsky et al., 1999)) and implementation quality (Lendrum, Humphrey, & Greenberg, in press).

A possible explanation for the inconsistent findings in the current study is the failure to account for the temporal aspect of programme development and evaluation. For instance, studies led by a developer may indicate an earlier or more formative stage of programme evaluation (see Campbell et al., 2000), where critical elements are still being trialled and modified. In this instance, it would be hypothesised that the 'involved' or 'independent' categories would begin to show greater effects, as the programme elements are finalised and the evaluations become more summative than formative (this would be conceptualised as increasing effects, similar to the pattern of results for SEC and CP in figure 2). However, this also suggests a limitation with the study methodology (specifically, a lack of independence between the two constructs of stage of evaluation and developer involvement). Also indicated is broader limitation with the current the status of the field. The majority of programme in the field are in relatively early stages of development and evaluation. (as the current findings show, approximately 69% are in the efficacy-based). Therefore interpretation of any other factors affecting programme success (e.g. developer involvement) are limited by the over-representation of this category. This limits the conclusions that can be drawn from beyond the preliminary exploration presented here.

This evidence does not preclude other hypotheses or future investigation of the potential effects of developer involvement in further research. Indeed, Eisner (2009) identifies a number of possible causal mechanisms to explain the general reduction of effects in independent prevention trials, and draws together a checklist by which consumers of research (and those researching these effects directly) can consider the extent to which these factors may influence results. Examples include cognitive biases, ideological position, and financial interests. Such an approach would aid clarity, as the further investigation of the phenomenon is currently limited by difficulty in establishing the precise role of the developer in individual trials when coding studies. To be able to examine the cynical / high fidelity views (see literature review) more thoroughly, studies need to more precisely report the extent and nature of the developer's involvement in an evaluation. Additional to this would be the consideration of implementation difficulties, as a significant minority of trials included in the present study did not report implementation. This would allow more comprehensive testing of the 'high fidelity' as a specific hypothesis beyond the cross tabulation analysis in table 7, which although not supportive of implementation quality as a moderator related to developer involvement, is a relatively blunt analysis (e.g. only containing studies which reported on implementation). Results from the current study tentatively suggest that the SEL field seems immune to the potential biases suggested by Eisner (2009), but there is little evidence to indicate why this would be so. Therefore, there is sufficient cause to further explore this issue, potentially using factors identified by Eisner as a starting point.

3) Studies implemented within the country of development will show larger effect sizes than those adopted and implemented outside country of origin

The hypothesis that programmes developed and implemented within their own national boundaries would show greater effects than those transported to another country was supported by 4 of the 7 outcomes (social-emotional competence, attitudes towards self, pro-social behaviour, emotional distress) all of these variables were statistically significant. Of particular note is the size of the

effects between categories, with some programmes showing almost no impact at all when transferred internationally.

Extant literature provides some guidance in explaining these effects. Several authors note the challenges associated with the implementation of programmes across cultural and international boundaries (Emshoff, 2008; Ferrer-Wreder, Sundell, & Mansoory, 2012; Resnicow, Soler, Braithwaite, Ahluwalia, & Butler, 2000), and therefore it is not surprising that the 'away' category would show reduced effects. For instance, the lack of critical infrastructure (e.g. quality supervision, site preparation, and staff training) has been an often cited reason for implementation failure (Elliott & Mihalic, 2004; Spoth et al., 2004). In this way, programmes may still be considered internally valid, but are not able to be implemented within a new context - flowers do not bloom in poor soil. This is a relatively optimistic interpretation of the data, as this would imply that the only limiting factor in successful implementation is a lack of more established 'ground work' in key areas (such as those identified by Elliott and Mihalic (2004), prior to the introduction of the programme. However, independent of infrastructure concerns, Kumpfer, Alvarado, Smith, and Bellamy (2002) note that interventions that are not aligned with the cultural values of the community in which they are implemented are likely to show a reduction in programme effects. This is consistent with earlier considerations in the literature. Wolf (1978) draws a distinction between the perceived need for programme outcomes (i.e. reduced bullying) and the procedures and processes for achieving this goal (i.e a specific taught curriculum). In practice, this would be consistent with the transportation of programmes which may be appropriate, but not congruent, with educational approaches or pedagogical styles. Contrary to the lack of infrastructure argument which requires school-based changes, it is programme adaptation that is required to address these needs. Berman and McLaughlin (1976) suggest that the likely answer is somewhere in the middle, with 'mutual adaptation' of both programme and implementation setting required for best results. Although these ideas are far from new, results from the current study suggest further understanding of the processes of cultural transportability of programmes is still undoubtedly required.

Neither infrastructure nor cultural adaptability fully explains why certain outcome variables show *larger* effect sizes in the 'away' category (i.e. why adapted programmes should outperform home versions). It may be that certain programmes require less external infrastructure, or may be more amenable to surface-level adaptations which do not interfere with change mechanisms and are therefore easier to transport. However, this would account for roughly equivalent, rather than enhanced, effects compared to home programmes. A partial explanation is offered when considering the temporal aspect of programme development alongside existing theories. For outcomes where larger effects are seen in the away category, it is possible that these programmes have an established history of development and evaluation in a broader range of contexts within the original country of development, resulting in greater external validity and subsequently fewer issues for transferability when exported. However, it is difficult to fully assess this hypothesis within the current design. This acknowledges a need for methodological diversity in in investigating these phenomena.

Following from this argument, the data may be representative of a much more cyclical (rather than linear) framework of design and evaluation as proposed by Campbell et al. (2000), by which large scale, 'successful' interventions are returned to a formative phase of development when implemented in a 'new' context with a new population (either within or across international boundaries). This is consistent with the idea of 'cultural tailoring' (Resnicow, et al., 2000), which is used to describe adaptations in interventions that are specifically targeted towards new cultural groups. Variation in programme outcomes may be representative of the ease and/or extent to which cultural tailoring requires re-validation of a programme, in line with Campbell et al.'s (2000) framework. In this way, the findings of the current study fail to fully capture this temporal, cyclical element, sampling from individual interventions at potentially each stage of development within its new cultural climate.

Limitations of the current study

The most pressing limitation of the current study is that of diversity, both in regards to 'methodological diversity' (variability in study design), and 'clinical heterogeneity' (differences between participants, interventions and outcomes) (the results of which are indicated by the Q and I^2 statistics (Higgins & Green, 2008)).   This suggests that the current categorisation of studies by the selected variables (trial stage, development involvement and cultural transportability) warrant further consideration in relation to their fit with the obtained data. i.e. they explain some, but not all, of the variability in outcomes.

The identified heterogeneity is in no small part due to the expansive definition by which SEL programmes are identified (Payton et al. 2008). This raises questions about the utility of such broad definitions within the academic arena as it currently precludes more precise investigations of specific issues.   For instance, the inconsistency by which relevant variables are directly related to an intervention's immediate or proximal outcome masks potential moderating effects when using meta-analytic strategies. As noted by Sklad et al. (2012), direct effects for some programmes (i.e. improved self-control) are considered as indirect by others (i.e. an intermediary part of a logic model in which pro-social behaviour is the intended outcome). This is an issue at theory level, contingent upon the logic models which underpin the implementation strategies of individual programme, but has serious implications for outcome assessment. For instance, lesser gains would be expected from distal outcomes, and therefore should not be assessed alongside proximal outcomes.

A further consideration is that the current study examined simple effects only; the potential moderating effects of stage of trial, development involvement, and cultural transportability as independent of one another. Table 3 demonstrates that the current field is not evenly balanced in relation to the factors, and therefore small cell sizes precluded the reliable examination of additive or interaction effects between these variables. This presents an intriguing avenue of enquiry beyond this preliminary investigation – e.g. To what extent could these factors inter-relate? Although theorised as independent constructs, there are hypothetical scenarios that would suggest these

factors can relate to one another, several of which have already been presented in this paper. For instance, we may hypothesise that there an interrelationship between trial stage and development involvement, when also considering the temporal aspect of programme development. However, additional theoretical work is required to further map the nature of these constructs and their relationships, as some combinations of factors are far less likely to originate in the field and may also prove counter-intuitive to established frameworks for programme development (e.g. Campbell et al, 2000). For instance, it is very unlikely to find a developer led effectiveness trail being delivered outside of its country of origin. Such considerations preclude 'simple' analyses such as cross tabulated frequencies, as these can be easily misinterpreted without further substantive theorisation and discussion.

It is argued that such theorisation should accompany further empirical work. For instance, future studies may consider the application of regression frameworks in which factors such as those already identified are used to predict outcomes, which may help address the overlap (or 'shared variance') between the constructs. This paper serves (in part) as a call for this kind further work in this area, (both theoretical and empirical). However, such approaches require further maturation and development within the field. As noted above, there are still relatively small numbers of programmes progressing to efficacy stage, and similarly very little cultural transportability. As more trials more into 'later stages' of development, more data and understanding will hopefully be forthcoming, furthering the basis of the preliminary exploration presented here.

The methodological limitations in the current study mirror the wider difficulties in the field; specifically the failure for some of the more commonly adopted methods to capture the complexities and nuances in relation to the unique ecology of the implementation context, most notably due to an emphasis on randomised control trial (RCT) methodology (and its variants), specifically incorporating 'Intent to Treat' (ITT) assessment (Gupta, 2011). It has been argued that RCT methodology is a limited approach, as trials can fail to explore how intervention components

and their mechanisms of change interact (Bonell, Fletcher, Morton, Lorenc, & Moore, 2012). For instance, how an intervention interacts with contextual adaptation, which has argued to be both inevitable and essential for a successful outcome (Ringwalt et al., 2003). This limitation is translated into the current study design, for instance it is worth considering the relatively blunt (yet effectively dichotomised) measure of cultural transferability used within the current study. Only international borders are considered, which does not take into account cultural variability *within* countries. In relation to the practice of cultural tailoring (Resnicow, et al., 2000), there is little methodological insight to help represent the diverse ecologies of classroom culture

In relation to the' ideal' amount of adaption for optimal outcomes, positive outcomes on certain measures are very much dependent on the context and composition of individual classes. For instance, although there is certainly evidence for *universal* gains for some outcomes (e.g. social-emotional competence), there are *differential* gains for others (e.g. pupils need to demonstrate poor behaviour before being able to improve on measures of conduct problems). Although there have been calls for RCTs comparing adapted versions of a programme to its 'generic' counterpart, (Kumpfer, Smith, & Bellamy, 2002), ITT and meta-analytic strategies (including the current study) are not optimally equipped to detect these subtleties (Oxman & Guyatt, 1992). An alternative 'middle-ground' is suggested by Hawe, Shiell and Riley (2004) regarding the need for flexibility in complex interventions. They suggest that for complex interventions (defined loosely as an intervention in which there is difficulty in precisely defining 'active ingredients' and how they inter-relate), it is the function and process that should be standardised and evaluated, not the components themselves (which are free to be adapted).

Future Directions and recommendations

The findings from the current study are evidence that the stakes continue to be high for the adoption of universal social and emotional learning (SEL) programmes. Although the field has firmly established that SEL can *potentially* be effective in addressing serious societal concerns of social-

emotional wellbeing and behaviour, there is comparatively limited understanding in how positive effects can be consistently maintained. As there is little caution in the speed of acceptance and roll out of SEL programmes internationally, despite these gaps in knowledge, findings of the current study have a global significance and present an opportunity to shape future directions and address several key lines of enquiry.

As SEL is a global phenomenon, the importance of additional work in understanding the significance of cultural validity specifically becomes increasingly important, given that results from the current study suggest that SEL programmes identified as successful can be rendered ineffective when transported to other countries. Aside from revising expectations of the likely effects that can be generated by an exported programme, there is arguably a wider methodological issue to be addressed when designing studies to assess transported programmes. For instance, additional work is needed in examining the potential importance of prior infrastructure across sites (such as those identified by Elliott and Mihalic (2004), and types and number of adaptations made (Berman & McLaughlin, 1976; Castro, Barrera, & Martinez, 2004; Hansen et al., 2013) occurs within a trial.

Addressing the recommendations above will require new thinking in methodological approaches in order to address the complexities of SEL interventions. The current study highlights both the strengths and weaknesses of meta-analytic approaches and therefore, a parallel but no less important, recommendation is for additional consideration of the complexity and heterogeneity of interventions using a full range of methodologies. Further meta-analytical approaches (e.g. by grouping studies into 'clinically meaningful units' (Melendez-Torres, Bonell, & Thomas, 2015) of function and process (Hawe, Shiell, & Riley, 2004) (e.g. mode of delivery)) alongside more 'bottom-up' approaches to examine the unique ecologies of individual classroom practices in more detail are advised.

There is an additional concern to better understand the internal logic models of individual programmes, ie. the active ingredients' and how they inter-relate. More clearly specifying the 'how'

and 'why' of programmes, would allow researchers to identify how various outcomes or 'ingredients' of SEL programmes are linked (Dirks, Treat, & Weersing, 2007). This is a daunting task, partially because difficulty in precisely defining 'active ingredients' is what makes an intervention complex. However, the methods employed should be guided by the substantive questions of the field. As literature is now addressing, not 'does SEL work?' (results have answered in the affirmative (Durlak et al., 2010; Sklad, 2012), but questions of '*how* does SEL work (or, why does it fail)?' new methodological thinking is required to answer these. This meta-analysis represents the first of many large steps required to address this question.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Achenbach, T. M., & Edelbrock, C. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington: University of Vermont Press.

*Anda, D. (1998). The evaluation of a stress management program for middle school adolescents. *Child and Adolescent Social Work Journal*, *15*, 73–85.

*Ando, M., Asakura, T., Ando, S., & Simons-Morton, B. (2007). A psychoeducational program to prevent aggressive behavior among Japanese early adolescents. *Health Education & Behavior 34*, 765–76. doi:10.1177/1090198106291965

*Ashdown, D. M., & Bernard, M. E. (2011). Can explicit instruction in Social and Emotional Learning skills Benefit the social-emotional development, well-being, and academic achievement of young children? *Early Childhood Education Journal*, *39*, 397–405. doi:10.1007/s10643-011-0481-x

*Baldry, A. C., & Farrington, D. P. (2004). Evaluation of an intervention program for the reduction of bullying and victimization in schools. *Aggressive Behavior*, *30*, 1–15. doi:10.1002/ab.20000

*Barrett, P., & Turner, C. (2001). Prevention of anxiety symptoms in primary school children: preliminary results from a universal school-based trial. *The British Journal of Clinical Psychology*, *40*, 399–410.

*Barrett, P. M., Farrell, L. J., Ollendick, T. H., & Dadds, M. (2006). Long-term outcomes of an Australian universal prevention trial of anxiety and depression symptoms in children and youth : An evaluation of the Friends program. *Journal of Clinical Child & Adolescent Psychology*, *35*, 403–411.

Beck, A., Steer, A., & Brown, K. (1996). *BDI-II Beck Depression Inventory, 2nd^{ed}*. San Antonio: The Psychological Corporation, Harcourt Brace& Company.

Beelmann, A., & Lösel, F. (2006). Child social skills training in developmental crime prevention: effects on antisocial behaviour and social competence. *Psychothema, 18*, 603-610

*Beets, M. W., Flay, B. R., Vuchinich, S., Snyder, F. J., Acock, A., Li, K.-K., … Durlak, J. (2009). Use of a social and character development program to prevent substance use, violent behaviors, and sexual activity among elementary-school students in Hawaii. *American Journal of Public Health*, *99*, 1438–45. doi:10.2105/AJPH.2008.142919

Berman, P., & McLaughlin, M. (1976). Implementation of educational innovations. *The Educational Forum*, *40*, 345-370.

*Bernard, M. E., & Walton, K. (2008). The effect of You Can Do It ! Education in six schools on student perceptions of well-being , teaching-learning and relationships. *Journal of Student Wellbeing, 5,* 1-23

*Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J., Greenberg, M. T., … Gill, S. (2008). Promoting academic and social-emotional school readiness: the head start REDI program. *Child Development*, *79*, 1802–17. doi:10.1111/j.1467-8624.2008.01227.x

*Bijstra, J. O., & Jackson, S. (1998). Social skills training with early adolescents: Effects on social skills, well-being, self-esteem and coping. *European Journal of Psychology of Education*, *13*, 569–583. doi:10.1007/BF03173106

Bodenman, G., Cina, A., Ledenmann, T., & Sanders, M. R. (2008). The efficacy of positive parenting program (Triple P) in improving parenting and child behaviour: A comparison with two other treatment conditions. *Behaviour Research and Therapy, 46*, 411–42.

*Bond, L., Patton, G., Glover, S., Carlin, J. B., Butler, H., Thomas, L., & Bowes, G. (2004). The Gatehouse Project: can a multilevel school intervention affect emotional wellbeing and health risk behaviours? *Journal of Epidemiology and Community Health*, *58*, 997–1003. doi:10.1136/jech.2003.009449

Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine, 45*, 2299-2306. doi:10.1016/j.socscimed.2012.08.032

*Bosworth, K. (2000). Preliminary evaluation of a multimedia violence prevention program for adolescents. *American Journal of Health Behaviour*, *24*, 268–280.

*Botvin, G. J., Griffin, K. W., & Nichols, T. D. (2006). Preventing youth violence and delinquency through a universal school-based prevention approach. *Prevention Science 7*, 403–8. doi:10.1007/s11121-006-0057-y

*Brackett, M. a., Rivers, S. E., Reyes, M. R., & Salovey, P. (2012). Enhancing academic performance and social and emotional competence with the RULER feeling words curriculum. *Learning and Individual Differences*, *22*, 218–224. doi:10.1016/j.lindif.2010.10.002

*Caldarella, P., Christensen, L., Kramer, T. J., & Kronmiller, K. (2009). Promoting Social and Emotional Learning in second grade Students: A study of the Strong Start curriculum. *Early Childhood Education Journal*, *37*, 51–56. doi:10.1007/s10643-009-0321-4

Campbell, M., Fitzpatrick, R., Haines, A., Kinmouth, A., Sandersock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, *321*, 694-696.

Castro F.G., Barrera, M., & Martinez, C.R. (2004). The cultural adaptation of prevention interventions: resolving tensions between fidelity and fit. *Prevention Science. 5*, 41-45.

*Catalano, R. F., Mazza, J. J., Harachi, T. W., Abbott, R. D., Haggerty, K. P., & Fleming, C. B. (2003). Raising healthy children through enhancing social development in elementary school: Results after 1.5 years. *Journal of School Psychology*, *41*, 143–164. doi:10.1016/S0022-4405(03)00031-1

Collaborative for Academic, Social and Emotional Learning. (2002). *Guidelines for social and emotional learning: High quality programmes for school and life success.* Chicago, IL : CASEL.

Committee for Children (2011). *Second Step: A Violence Prevention Curriculum*. Seattle, WA: Committee for Children.

*Conduct Problems Prevention Research Group. (1999). Intial impact of the fast track prevention trial for conduct problems: I The high risk sample. *Journal of Consulting and Clinical Psychology*, *67*, 631–647.

Cooke, M., Ford, J., Levine, J., Bourke, C., Newell, L., & Lapidus, G. (2007). The effects of city-wide implementation of "Second Step" on elementary school students' prosocial and aggressive behaviors. *The Journal of Primary Prevention*. *28*, 93–115.

*Cunningham, E. G., Brandon, C. M., & Frydenberg, E. (2002). Enhancing coping resources in early adolescence through a school-based program teaching optimistic thinking skills. *Anxiety, Stress & Coping*, *15*, 369–381. doi:10.1080/1061580021000056528

*Curtis, C., & Norgate, R. (2007). An evaluation of the Promoting Alternative Thinking Strategies curriculum at Key Stage 1. *Educational Psychology in Practice*, *23*, 33–44. doi:10.1080/02667360601154717

Dane, A.V. & Schneider, B.H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23-45.

Denham, S. A. (2005). *Assessing Social-Emotional Development in Children From a Longitudinal Perspective for the National Children's Study*. Columbus, OH: Battelle Memorial Institute.

Department for Education and Skills. (2007). *Social and Emotional Aspects of Learning for Secondary Schools (SEAL) Guidance Booklet.* London*:* DfES.

Dersimonian, R., & Laird, N. (1986). Meta-Analysis in clinical trials. *Controlled Clinical Trials, 7*, 177-188.

Dunn, L.,& Dunn, L.(1997). *Peabody Picture Vocabulary Test – Revised*. Circle Pines, MN: American Guidance Service.

*Durant, R. H., Barkin, S., & Krowchuk, D. P. (2001). Evaluation of a peaceful conflict resolution and violence prevention curriculum for sixth-grade students. *The Journal of Adolescent Health : Official Publication of the Society for Adolescent Medicine*, *28*, 386–93.

*Durant, R. H., Treiber, F., Getts, A., McCloud, K., Linder, C. W., & Woods, E. R. (1996). Comparison of two violence prevention curricula for middle school adolescents. *The Journal of Adolescent Health : Official Publication of the Society for Adolescent Medicine*, *19*, 111–7. doi:10.1016/1054-139X(96)00030-4

Durlak, J.A. (1998). Why programme implementation is important. *Journal of Prevention and Intervention in the Community. 17*, 5-18.

Durlak, J., & DuPre, E. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Journal of Community Psychology. 41*, 327–350.

Durlak, J., Weissberg, R., Dymicki, A., Taylor, R., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*, 405–432.

Dusenbury, L., Brannigan, R., Falco, M. & Hansen, W.B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings, *Health and Education Research. 18*, 237-256.

Eisner, M. (2009). No effects in independent prevention trials: can we reject the cynical view? *Journal of Experimental Criminology. 5,* 163-183.

Elias, M.J., Lantieri, L., Patti, J., Shriver, T.P., Walberg, H. Weissberg, R.P. and Zins, J.E. (2001). No new wars needed. *Reclaiming Children and Youth, Spring*, 29-32.

Elliott, J., & Kushner, S. (2007). The need for a manifesto for educational programme evaluation. Cambridge Journal of Education, *37*, 321-336. doi: 10.1080/03057640701546649

Elliott, D., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programmes. *Prevention Science, 5*, 47-52.

Emshoff, J., (2008), Researchers, practitioners, and funders: Using the framework to get us on the same page. *American Journal of Community Psychology, 41*, 393-403. doi: 10.1007/s10464-008-9168-x

*Essau, C., Conradt, J., Sasagawa, S., & Ollendick, T. H. (2012). Prevention of anxiety symptoms in children: results from a universal school-based trial. *Behavior Therapy*, *43*, 450–64. doi:10.1016/j.beth.2011.08.003

Evans, J., & Benefield, P. (2001). Systematic reviews of educational research: Does the medical model fit? British Educational Research Journal, *27*, 527-541. doi/10.1080/01411920120095717

*Farrell,A. D., & Meyer,A. L. (1997). The effectiveness of a school-based curriculum for reducing violence among urban sixth-grade students. *American Journal of Public Health*, *87*, 979–84.

*Farrell, A. D., Meyer, A. L., & White, K. S. (2001). Evaluation of Responding in Peaceful and Positive Ways (RIPP): A school-based prevention program for reducing violence among urban adolescents. *Journal of Clinical Child & Adolescent Psychology*, *30*, 451–463.

*Farrell, A. D., Meyer, A. L., Sullivan, T. N., & Kung, E. M. (2003). Evaluation of the Responding in Peaceful and Positive Ways (RIPP) seventh grade violence prevention curriculum. *Journal of Child and Family Studies*, *12*, 101–120.

*Farrell, A. D., Valois, R. F., Meyer, A. L., & Tidwell, R. P. (2003). Impact of the RIPP Violence Prevention Program on rural middle school students. *The Journal of Primary Prevention*, *24*, 143–167.

*Fekkes, M., Pijpers, F. I. M., & Verloove-Vanhorick, S. P. (2006). Effects of antibullying school program on bullying and health complaints. *Archives of Pediatrics & Adolescent Medicine*, *160*, 638–44. doi:10.1001/archpedi.160.6.638

Ferrer-Wreder, L., Adamson, L., Kumpfer, K, & Eichas, K. (2012). Advancing intervention science through effectiveness research: A global persective. Child Youth Care Forum, *41*, 109-117. doi: 10.1007/10566-012-9173-y

Ferrer-Wreder, L., Sundell, K., & Mansoory, S. (1012). Tinkering with perfection: Theory development in the intervention cultural adaptation field. *Child Youth Care Forum, 41*, 149-171. doi: 10.1007/s10566-011-9162-6

*Flannery, D. J., Vazsonyi, A. T., Liau, A. K., Guo, S., Powell, K. E., Atha, H., … Embry, D. (2003). Initial behavior outcomes for the PeaceBuilders universal school-based violence prevention program. *Developmental Psychology*, *39*, 292–308. doi:10.1037/0012-1649.39.2.292

Flay, B.R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine, 15,* 451-474

*Flay, B. R., Allred, C. G., & Ordway, N. (2001). Effects of the Positive Action program on achievement and discipline: two matched-control comparisons. *Prevention Science*, *2*, 71–89.

Flay, B. R., Biglan, A., Boruch, F. R., Castro, G. F., Gottfredson, D., Kellam, K., Moscicki, K. E., Schinke, S., Valentine, C. J., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemintation. Prevention Science, *6*, 151-175. doi: 10.1007/s11121-005-5553-y

Forman, S., Shapiro, E. Codding, R. & Gonzales, J. (2013). Implementation science and school psychology. *School Psychology Quarterly, 28,* 77-100.

*Frey, K. S., Nolen, S. B., Van Schoiack Edstrom, L., & Hirschstein, M. K. (2005). Effects of a school-based social–emotional competence program: Linking children's goals, attributions, and behavior. *Journal of Applied Developmental Psychology*, *26*, 171–200. doi:10.1016/j.appdev.2004.12.002

*Garaigordobil, M., & Echebarria, A. (1995). Assessment of a peer-helping game program on children's development. *Journal of Research in Childhood Education*, *10*, 63–69.

*Gillham, J. E., Reivich, K. J., Freres, D. R., Chaplin, T. M., Shatté, A. J., Samuels, B., … Seligman, M. E. P. (2007). School-based prevention of depressive symptoms: A randomized controlled study of the effectiveness and specificity of the Penn Resiliency Program. *Journal of Consulting and Clinical Psychology*, *75*, 9–19. doi:10.1037/0022-006X.75.1.9

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child Adolescent Psychiatry, 40*, 1337–1345

*Goossens, F., Gooren, E., Orobio de Castro, B., Overveld, K., Buijs, G., Monshouwer, K., … Paulussen, T. (2012). Implementation of PATHS through Dutch municipal health services : A quasi-experiment *International Journal of Conflict and Violence*, *6*, 234–248.

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. *The International Journal of Mental Health Promotion*, *10*, 13–20.

Greenberg, M. T. (2010). School-based prevention: Current status and future challenges. *Effective Education. 2,* 27–52.

Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A. & Zins, J. E. (2005). *The study of implementation in school-based preventive interventions: Theory, practice and research*.   Washington, DC: Department of Health and Human Services.

Greenberg, M., & Kusché, C. (2002). *Promoting Alternative Thinking Strategies (PATHS), blueprints for violence prevention*. Boulder: Center for the Study and Prevention of Violence, University of Colorado.

*Greenberg, M. T., Kusche, C., Cook, E. T., & Quamma, J. P. (1995). Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. *Development and Psychopathology*, *7*, 117. doi:10.1017/S0954579400006374

Greenberg, M., Weissberg, R., O'Brien, M., Zins, J., Fredericks, L., Resnik, H., and Elias, M. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist. 58*, 466–474.

Gresham, F., & Elliot, S. (1990). *Social Skills Rating System*. Circle Pines: American Guidance Service.

*Gueldner, B., & Merrell, K. (2011). Evaluation of a Social-Emotional Learning program in conjunction With the wxploratory application of performance feedback incorporating motivational interviewing techniques. *Journal of Educational and Psychological Consultation*, *21*, 1–27. doi:10.1080/10474412.2010.522876

Gupta, S. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research, 2*, 109-112.

*Haines, J., Neumark-Sztainer, D., Perry, C. L., Hannan, P. J., & Levine, M. P. (2006). V.I.K. (Very Important Kids): a school-based program designed to reduce teasing and unhealthy weight-control behaviors. *Health Education Research*, *21*, 884–95. doi:10.1093/her/cyl123

Hansen, W., Pankratz, M., Dusenbury, L., Giles, S., Bishop, D., Albritton, J., Albritton, L., & Strack, J. (2013). Styles of adaptation: The impact of frequency and valence of adaptation on preventing substance use. *Health Education, 113*, 345-363. doi: http://dx.doi.org/10.1108/09654281311329268

*Harlacher, J. E., & Merrell, K. W. (2010). Social and Emotional Learning as a universal level of student support: Evaluating the follow-up effect of strong Kids on social and emotional outcomes. *Journal of Applied School Psychology*, *26*, 212–229. doi:10.1080/15377903.2010.495903

Hallfors, D. & Godette, D. (2002). Will the 'Principles of Effectiveness' improve prevention practice? Early findings from a diffusion study. *Health Education Research. 17*, 461-470.

*Harrington, N., Giles, S., Hoyle, R., Feeney, G., & Yungbluth, S. (2001). Evaluation of the All Stars character education and problem behavior prevention program: effects on mediator and outcome variables for middle school students. *Health Education and Behaviour*, *28*, 533–546.

Hawe, P., Shiell, A., Riley, T. (2004). Complex interventions: How "out of control" cana radonised control trial be? *British Medical Journal, 238*, 1561-1563. doi: http://dx.doi.org/10.1136/bmj.328.7455.1561

Hedges, L. & Olkin, I. (1985). *Statistical Methods for meta-analysis*. New York: Academic Press.

Hedges, L., & Pigott, T. (2005). The power of statistical tests in meta-analysis. *Psychological Methods, 6*, 203-217.

*Hennessey, B. (2007). Promoting social competence in school-aged children: The effects of the open circle program. *Journal of School Psychology*, *45*, 349–360. doi:10.1016/j.jsp.2006.11.007

Henningham, H. (2013). *Transporting an evidence-based programme to Jamacian preschools: Cluster randomised controlled trial*. Paper presented at the 21st society for prevention research conference, San Francisco.

Higgins, J., & Green, S. (2008). *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, UK: John Wiley & Sons.

Higgins, J., Thompson, S., Deeks, J., & Altman, D. (2003). Measuring inconsistency in meta-analyses. *BMJ 327*, 557–60. doi:10.1136/bmj.327.7414.557

*Holen, S., Waaktaar, T., & Health, A. M. (2012). The effectiveness of a universal school- based programme on coping and mental health : a randomised control study of Zippy's Friends. *Educational Psychology, 32*, 657–677.

*Holsen, I., Smith, B. H., & Frey, K. S. (2008). Outcomes of the social competence program Second Step in Norwegian elementary schools. *School Psychology International*, *29*, 71–88. doi:10.1177/0143034307088504

*Horowitz, J. L., Garber, J., Ciesla, J., Young, J. F., & Mufson, L. (2007). Prevention of depressive symptoms in adolescents: a randomized trial of cognitive-behavioral and interpersonal prevention programs. *Journal of Consulting and Clinical Psychology*, *75*, 693–706. doi:10.1037/0022-006X.75.5.693

*Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence. *Journal of Emotional and Behavioral Disorders*, *9*, 146–160.

Jinks, J., & Morgan, V. (1999). Children's perceived academic self-efficacy: An inventory scale. *The Clearing House: A Journal of Educational Research, Controversy, and Practices, 72*, 224-230.

*Jones, S. M., Brown, J. L., Hoglund, W. L. G., & Aber, J. L. (2010). A school-randomized clinical trial of an integrated social-emotional learning and literacy intervention: impacts after 1 school year. *Journal of Consulting and Clinical Psychology*, *78*, 829–42. doi:10.1037/a0021383

*Karasimopoulou, S., Derri, V., & Zervoudaki, E. (2012). Children's perceptions about their health-related quality of life: effects of a health education-social skills program. *Health Education Research*, *27*, 780–93. doi:10.1093/her/cys089

*Kimber, B., Sandell, R., & Bremberg, S. (2008). Social and emotional training in Swedish schools for the promotion of mental health: an effectiveness study of 5 years of intervention. *Health Education Research*, *23*, 931–40. doi:10.1093/her/cyn040

Kontopantelis, E., & Reeves D. (2009). MetaEasy: A meta-analysis add-in for Microsoft Excel. *Journal of Statistical Software,30*, 1-25.

Kumpfer, K., Alvarado, R., Smith, P., & Bellamy, N. (2002). Cultural sensitivity and adaptation in family-based prevention interventions. Prevention Science, *3,* 241- 246. doi: 10.1023/A:1019902902119

Kumpfer, K. L., Magalhães, C. & Xie, J. (2012). Cultural adaptations of evidence-based family interventions to strengthen families and improve children's developmental outcomes. *European Journal of Developmental Psychology, 9*, 104-116.

*Lau, N., & Hue, M. (2011). Preliminary outcomes of a mindfulness-based programme for Hong Kong adolescents in schools : well-being, stress and depressive symptoms. *International Journal of Children's Spirituality*, *16*, 315–330.

*Leadbeater, B., Hoglund, W., & Woods, T. (2003). Changing contexts? The effects of a primary prevention program on classroom levels of peer relational and physical victimization. *Journal of Community Psychology*, *31*, 397–418. doi:10.1002/jcop.10057

*Leming, J. S. (2000). Tell Me a Story : An evaluation education programme. *Journal of Moral Education*, *29*, 413–427. doi:10.1080/03057240020015029

Lendrum, A., Humphrey, N. & Greenberg, M.T. (in press). Implementing for success in school-based mental health promotion: The role of quality in resolving the tension between fidelity and adaptation. In Shute, R. & Slee, P. (Eds.). *Mental health and wellbeing through schools: The way forward*. Hove: Routledge

Lendrum, A., & Wigelsworth, M. (2013). The evaluation of school-based social and emotional learning interventions: Current issues and future directions. *Psychology of Education Review, 37*, 70-74.

*Linares, L. O., Rosbruch, N., Stern, M. B., Edwards, M. E., Walker, G., Abikoff, H. B., & Alvir, J. M. J. (2005). Developing cognitive-social-emotional competencies to enhance academic learning. *Psychology in the Schools*, *42*, 405–417. doi:10.1002/pits.20066

Little, M., Berry, V., Morpeth, L., Blower, S., Axford, N., Taylor, R., Bywater, T., Lehtonen, M., & Tobin, K. (2012). The impact of three evidence-based programmes delivered in public systems in Birmingham, UK. *International Journal of Conflict and Violence*, *6*, 260-272.

Lowry-Webster, H. M., Barrett, P. M., & Dadds, M. R. (2001). A universal prevention trial of anxiety and depressive symptomatology in childhood: Preliminary data from an Australian study. *Behaviour Change*, *18*, 36–50. doi:10.1375/bech.18.1.36

Luborsky, L., Diguer, L., Seligman, D., Rosenthal, R., Krause, E., Johnson, S., et al. (1999). The researcher's own therapy allegiances: a "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*, 95-106

*Malti, T., Ribeaud, D., & Eisner, M. (2012). Effectiveness of a universal school-based social competence program : The Role of child characteristics and economic factors. *International Journal of Conflict and Violence*, *6*, 249–259.

*McClowry, S. G., Snow, D. L., & Tamis-LeMonda, C. S. (2005). An evaluation of the effects of INSIGHTS on the behavior of inner city primary school children. *The Journal of Primary Prevention*, *26*, 567–84. doi:10.1007/s10935-005-0015-7

*McClowry, S. G., Snow, D. L., Tamis-Lemonda, C. S., & Rodriguez, E. T. (2010). Testing the efficacy of INSIGHTS on student disruptive behavior, classroom management, and student competence in inner city primary grades. *School Mental Health*, *2*, 23–35. doi:10.1007/s12310-009-9023-8

Melendez-Torres, G., Bonell, C., & Thomas, J. (2015). Emergent approaches to the meta-analysis of multiple heterogeneous complex interventions. *BMC Research Methodology*, *15*, 47. doi: 10.1186/s12874-015-0040-z

*Mendelson, T., Greenberg, M. T., Dariotis, J. K., Gould, L. F., Rhoades, B. L., & Leaf, P. J. (2010). Feasibility and preliminary outcomes of a school-based mindfulness intervention for urban youth. *Journal of Abnormal Child Psychology*, *38*, 985–94. doi:10.1007/s10802-010-9418-x

*Merry, S., Wild, C. J., Bir, J., & Cunliffe, R. (2004). A Randomized placebo-controlled trial of a school-based depression prevention program. *Journal of the American Academy of Children and Adolescent Psychiatry*, *43*, 538-547. doi:10.1097/01.chi.0000117063.63530.68

*Meyer, G., Roberto, A. J., Boster, F. J., & Roberto, H. L. (2004). Assessing the Get Real about violence curriculum : Process and outcome evaluation results and implications. *Health Communication*, *16*, 451–474.

*Mikami, A. Y., Boucher, M., & Humphreys, K. (2005). Prevention of peer rejection through a classroom-level intervention in middle school. *The Journal of Primary Prevention*, *26*, 5–23. doi:10.1007/s10935-004-0988-7

*Mishara, B. L., & Ystgaard, M. (2006). Effectiveness of a mental health promotion program to improve coping skills in young children: Zippy's Friends. *Early Childhood Research Quarterly*, *21*, 110–123. doi:10.1016/j.ecresq.2006.01.002

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery, 8*, 226-241.doi: 10.1016/j.ijsu.2010.02.007

*Multisite, T., & Prevention, V. (2009). The ecological effects of universal and selective violence prevention programs for middle school students: a randomized trial. *Journal of Consulting and Clinical Psychology*, *77*, 526–42. doi:10.1037/a0014395

*Munoz, M., & Vanderhaar, J. (2006). Literacy-embedded character education in a large urban district: Effects of the child development project on elementary school students and teachers. *Journal of Research in Character Education*, *4*, 47–64.

*Orpinas, P., Kelder, S., Frankowski, R., Murray, N., Zhang, Q., & McAlister, A. (2000). Outcome evaluation of a multi-component violence-prevention program for middle schools: the Students for Peace project. *Health Education Research*, *15*, 45–58.

Oxman, A., & Guyatt, G. (1992). A consumer's guide to subgroup analyses. *Annals of Internal Medicine, 116*, 78-84.

*Patton, G. C., Bond, L., Carlin, J. B., Thomas, L., Butler, H., Glover, S., … Bowes, G. (2006). Promoting social inclusion in schools: a group-randomized trial of effects on student health risk behavior and well-being. *American Journal of Public Health*, *96*, 1582–7. doi:10.2105/AJPH.2004.047399

Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008). *The Positive Impact of Social and Emotional Learning for Kindergarten to Eighth-Grade Students*. Chicago: CASEL.

Perlis, R. H., Perlis, C. S., Wu, Y., Hwang, C., Joseph, M., & Nierenberg, A. A. (2005). Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *American Journal of Psychiatry, 162,* 1957–1960.

*Petermann, F., & Natzke, H. (2008). Preliminary results of a comprehensive approach to prevent antisocial behaviour in preschool and primary school pupils in Luxembourg. *School Psychology International*, *29*, 606–626. doi:10.1177/0143034308099204

Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology, 1*, 435-450.

*Pickens, J. (2009). Socio-emotional programme promotes positive behaviour in preschoolers. *Child Care in Practice*, *15*, 261–278. doi:10.1080/13575270903149323

*Raimundo, & Marques-Pinto. (2013). The effects of a social-emotional e-learning program on elementary school children: The role of pupil characteristics. *Psychology in the Schools, 50,* 165-180*. Doi*: 10.1002/pits.21667.

Rescicow, K., Soler, R., Braithwaite, R., Ahluwalia, J., & Butler, J. (2000). Cultural sensitivity in substance use prevention. *Journal of Community Psychology, 28*, 271-290.

*Riggs, N. R., Greenberg, M. T., Kusché, C., & Pentz, M. A. (2006). The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: effects of the PATHS Curriculum. *Prevention Science*, *7*, 91–102. doi:10.1007/s11121-005-0022-1

Ringwalt, C.L., Ennett, S., Johnson, R., Rohrbach, L.A., Simons-Rudolph, A., Vincus, A. & Thorne, J. (2003). Factors associated with fidelity to substance use prevention curriculum guides in the nation's middle schools. *Health Education Behavior, 30,* 375-391.

*Roberts, L. (2004). Theory development and evaluation of project WIN: A violence reduction program for early adolescents. *The Journal of Early Adolescence*, *24*, 460–483. doi:10.1177/0272431604268547

*Sanz de Acedo Lizarraga, M. L., Ugarte, M. D., Cardelle-Elawar, M., Iriarte, M. D., & Sanz de Acedo Baquedano, M. T. (2003). Enhancement of self-regulation, assertiveness, and empathy. *Learning and Instruction*, *13*, 423–439. doi:10.1016/S0959-4752(02)00026-9

*Sawyer, M. G., MacMullin, C., Graetz, B., Said, J., Clark, J. J., & Baghurst, P. (1997). Social skills training for primary school children: a 1-year follow-up study. *Journal of Paediatrics and Child Health*, *33*, 378–83.

*Schoiack-Edstrom, Frey, & Beland. (2008). Changing adolsecents' attitudes about relational and physical aggression: An early evaluation of a school-based intervention. *School Psychology Review, 31, 201-216.*

*Schonert-Reichl, K. a., & Lawlor, M. S. (2010). The Effects of a mindfulness-based education program on pre- and early Adolescents' well-being and social and emotional competence. *Mindfulness*, *1*, 137–151. doi:10.1007/s12671-010-0011-8

*Schonert-Reichl, K., Smith, V., Zaidman-Zait, A., & Hertzman, C. (2011). Promoting children's prosocial behaviors in school: Impact of the "Roots of Empathy" program on the social and emotional competence of school-aged children. *School Mental Health*, *4*, 1–21. doi:10.1007/s12310-011-9064-7

*Seifer, R., Gouley, K., Miller, A. L., & Zakriski, A. (2010). Early education & development implementation of the PATHS curriculum in an urban elementary school. *Early Education and Development,* 15, 471-486. Doi: 10.1207/s15566935eed1504_6

Shields, A., & Cicchetti, D. (1997). Emotion regulation among school-age children: The development and validation of a new criterion Q-sort scale. *Developmental Psychology, 33*, 906-916.

*Shochet, I. M., Dadds, M. R., Holland, D., Whitefield, K., Harnett, P. H.., & Osgarby, S. M. (2001). The efficacy of a universal school-based program to prevent adolescent depression. *Journal of Clinical Child & Adolescent Psychology*, *30*, 303–315.

Sklad, M., Diekstra, R., Ritter, M., & Ben, J. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and. *Psychology in the Schools, 49*, 892–909. doi:10.1002/pits

Social and Character Development Research Consortium (2010). *Efficacy of Schoolwide Programs to Promote Social and Character Development and Reduce Problem Behavior in Elementary School Children* (NCER 2011-2001). Washington, DC: U.S. Government Printing Office.

*Solomon, D., Battistich, V., Watson, M., & Lewis, C. (2000). A six-district study of educational change : direct and mediated effects of the child development project. *Social Psychology of Education, 4,* 3–51.

*Sørlie, M., & Ogden, T. (2007). Immediate Impacts of PALS: A school-wide multi-level programme targeting behaviour problems in elementary school. *Scandinavian Journal of Educational Research*, *51*, 471–492. doi:10.1080/00313830701576581

*Spence, S. H., Sheffield, J. K., & Donovan, C. L. (2003). Preventing adolescent depression: An evaluation of the Problem Solving For Life program. *Journal of Consulting and Clinical Psychology*, *71*, 3–13. doi:10.1037/0022-006X.71.1.3

Spoth, R., Greenberg, M., Bierman, K., & Redmond, C. (2004). PROSPER community-university partnership model for public education systems: Capacity-building for evidence-based, competence-building prevention. *Prevention Science, 5*, 31-39.

Stallard, P., Taylor, G., Anderson, R., Daniels, H., Simpson, N., Phillips, R., & Skryabina, E. (2014). The prevention of anxiety in children through school-based interventions: study protocol for a 24-month follow-up of the PACES project. *Trials, 15*, 77. doi:10.1186/1745-6215-15-77

*Stevahn, L., Johnson, D. W., Johnson, R. T., Green, K., & Laginski, A. M. (1997). Effects on high school students of conflict resolution training integrated into English literature. *The Journal of Social Psychology*, *137*, 302–315. doi:10.1080/00224549709595442

*Stevahn, L., Johnson, D. W., Johnson, R. T., & Real, D. (1996). The impact of a cooperative or individualistic context on the effectiveness of conflict resolution training. *American Educational Research Journal*, *33*, 801–823.

*Stevahn, L., Johnson, D. W., Roger, T., & Schultz, R. (2002). Effects of conflict Resolution training integrated Into a high school social studies curriculum. *The Journal of Social Psychology*, *142*, 305–331.

Stoolmiller, M., Eddy, J. M., Reid, J. B., & Social, O. (2000). Detecting and describing preventive intervention effects in a universal school-based randomized trial targeting delinquent and violent behavior pre score. Journal of Consulting and Clinical Psychology*, *68*, 296–306.

St.Pierre, T.L., Osgood, D.W., Mincemoyer, C.C., Kaltreider, D.L. & Kauh, T.J. (2005). Results of an independent evaluation of Project ALERT delivered in schools by cooperative extension. *Prevention Science, 6,* 305-317.

*Taub, J. (2001). Evaluation of the Second Step violence prevention program at a rural elementary School. *School Psychology Review*, *31*, 186–200.

*Teglasi, H., & Rothman, L. (2001). STORIES: A classroom-based program to reduce aggressive behavior. *Journal of School Psychology*, *39*, 71–94. doi:10.1016/S0022-4405(00)00060-1

Ttofi, M.M., Farrington, D. & Baldry, A.C. (2008). *Effectiveness of Programmes to Reduce School Bullying: A systematic review of school bullying programmes.* Stockholm: Swedish National Council for Crime Prevention.

*Twemlow, S., Fonagy, P., Sacco, F., Gies, M., Evans, R., & Ewbank, R. (2001). Creating a peaceful school learning environment: A controlled study of an elementary school intervention to reduce violence. *American Journal of Psychiatry, 158*, 808–810.

Webster-Stratton, C. (2011). *The Incredible Years: Parents, Teachers, and Children's Training Series*. Seattle: Incredible Years Inc.

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,* 29-69. doi: 10.1093/heapro/dar075.

Wehby, J., Maggin, D., Partin, T., & Robertson, R. (2011). The impact of working alliance, social validity, and teacher burnout on implementation fidelity of the Good Behaviour Game. *School Mental Health, 4,* 22-33

*Wigelsworth, M., Humphrey, N., & Lendrum, A. (2011). A national evaluation of the impact of the secondary Social and Emotional Aspects of Learning (SEAL) programme. *Educational Psychology*, *1*, 213–238.

Wiggins M., Bonnell C., Sawtell M., Austerberry H., Burchett H., Allen E. & Strange V. (2009). Health outcomes of youth development programmes in England: prospective matched comparison study. *British Medical Journal*, *339*, 2543.

Wilson, S. J., & Lipsey, M. W. (2007). School-Based Interventions for Aggressive and Disruptive Behavior. Update of a Meta-Analysis. *American Journal of Preventative Medicine*, *3*, 130–143.

*Witt, C., Becker, M., Bandelin, K., Soellner, R., Ph, D., & Willich, S. N. (2005). Qigong for Schoolchildren : A Pilot Study. *The Journal of Alternative and Complementary Medicine*, *11*, 41–47.

Wolf, M. (1978). Social validity: The case for subjective measurement. *Journal of Applied Behaviour Analysis, 11*, 203-214.

Zins, J., Bloodworth, M., Weissberg, R., & Walberg, H. (2007). The Scientific Base Linking Social and Emotional Learning to School Success. *Journal of Educational and Psychological Consultation, 17*, 191-210.

Tables and Figures

Table 1. *Denham's framework of social and emotional competence*

Emotional competence skills	Self-awareness	Understanding self-emotions
	Self-management	Emotional and behavioural regulation
	Social awareness	Understanding emotions
		Empathy/sympathy
Relational/pro-social skills	Social problem solving	Co-operation
		Listening skills
		Turn-taking
		Seeking help

Adapted from Denham (2005)

Table 2. *Study Characteristics*

Study Characteristic	N	%
Time of Assessment		
- Post-test (0 to 6 months)	63	71
- Follow up (6 to 18 months)	19	21
- Follow up (18+ months)	7	8
Period of Schooling (Based on USA school system)		
- Pre-school (4-6yrs)	7	8
- Elementary (7-11yrs)	63	71
- Middle (12-14yrs)	13	15
- High (15-18yrs)	6	6
Type of Design		
- RCT	14	16
- Cluster RCT	43	48
- wait-list (non-randomised)	2	2
- Matched Comparison	22	25
- Other	8	9
Implementation reported		
-Yes	61	69
- No	28	31
Of those reporting yes:		
Issues with Implementation		
- Yes	22	36
- No	39	64

Table 3. *Study Characteristics relating to specific hypotheses*

Study Characteristic	N	%
Efficacy / Effectiveness		
- Efficacy	61	69
- Effectiveness	25	28
- Unknown	3	3
Developer Involvement		
- Lead	34	38
- Involved	25	28
- Independent	30	34
Home / Away		
- Home	71	80
- Away	18	20

Table 4. *Mean effects, confidence intervals and fit statistics for main programme effects*

Outcome	N	Effect (95% CI)	Q	I^2
Social-emotional competence (SEC)	24	**0.53** (0.32-0.75)	832.36*	97%
Attitudes towards self (ATS)	9	**0.17** (0.07-0.28)	12.19	43%
Pro-social behaviour (PSB)	39	**0.33** (0.24-0.42)	362.35*	90%
Conduct problems (CP)	40	**0.28** (0.20-0.36)	411.44*	91%
Emotional distress (ED)	32	**0.19** (0.13-0.25)	101.15*	69%
Academic achievement (AA)	15	**0.28** (0.18-0.40)	40.63*	66%
Emotional competence only (ECO)	14	**0.27** (0.14-0.39)	72.63*	82%

Table 5. *Mean effects, confidence intervals and fit statistics for stage of evaluation (efficacy vs. effectiveness)*

Outcome		Efficacy				Effectiveness		
	n	Effect (95% CI)	Q	I^2 (%)	n	Effect (95% CI)	Q	I^2 (%)
SEC	14	**0.31** (0.20-0.42)	28.74*	58	10	**0.47** (0.18-0.76)	368.19*	98
ATS	4	**0.21** (-0.18-0.45)	6.59	55	5	**0.15** (0.05-0.26)	5.55	28
PSB	28	**0.37** (0.25-0.50)	231.84*	89	12	**0.18** (0.08-0.28)	61.94*	82
CP	26	**0.34** (0.20-0.48)	353.71*	93	13	**0.15** (0.07-0.22)	39.12*	69
ED	27	**0.24** (0.14-0.34)	210.61*	87	6	**0.14** (0.08-0.20)	5.26	5
AA	11	**0.38** (0.22-0.54)	31.87*	69	5	**0.22** (0.05-0.39)	14.54*	73
ECO	9	**0.28** (0.15-0.40)	22.43*	64	5	**0.23** (-0.02-0.48)	46.12*	91

Figure 1. *Effect size and 95% confidence intervals for each of the seven outcome variables for studies coded as either efficacy or effectiveness trials*
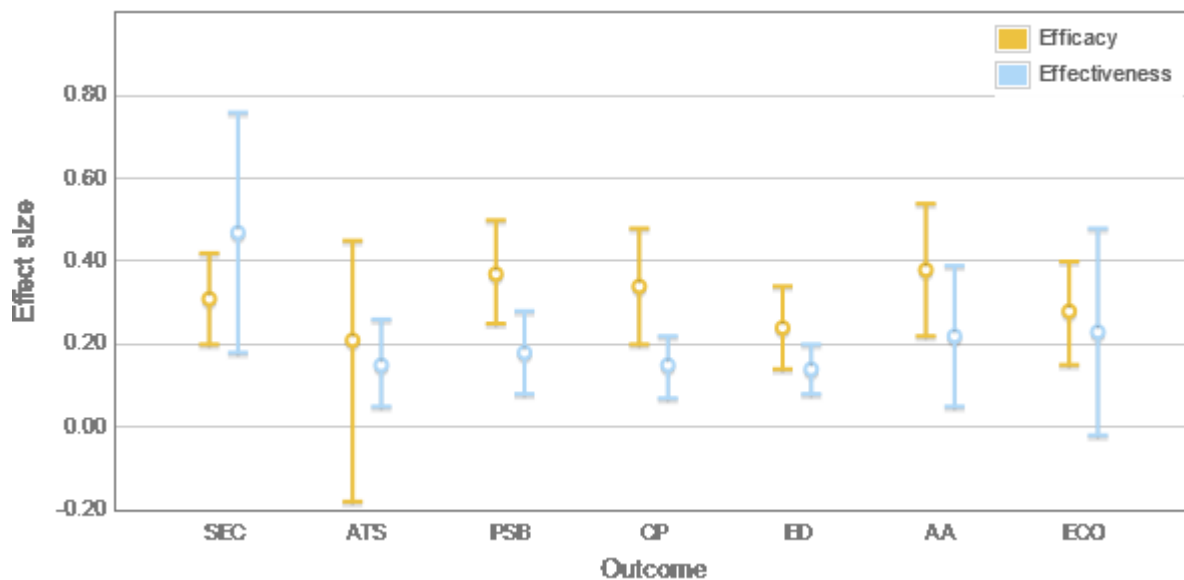
Table 6. *Mean effects, confidence intervals fit statistics for developer involvement (led vs. involved vs. independent)*

Outcome	Led				Involved				Independent			
	n	Effect (95% CI)	Q	I^2 (%)	n	Effect (95% CI)	Q	I^2 (%)	n	Effect (95% CI)	Q	I^2 (%)
SEC	5	**0.21** (0.04-0.39)	10.40*	62	7	**0.48** (0.23-0.73)	29.55*	80	12	**0.69** (0.33-1.05)	780.12*	99
ATS	5	**0.22** (0.08-0.37)	6.51	39	2	**0.02** (-0.16-0.54)	3.11	68	2	-0.003 (0.02-0.20)	0.01	-
PSB	10	**0.25** (0.03-0.46)	66.58*	86	14	**0.43** (0.26-0.60)	125.94*	90	15	**0.29** (0.17-0.42)	66.58*	86
CP	15	**0.20** (0.12-0.28)	36.26*	64	12	**0.25** (0.12-0.38)	61.60*	82	14	**0.37** (0.18-0.57)	311.82*	96
ED	18	**0.21** (0.13-0.29)	29.02*	45	6	**0.14** (0.00-0.26)	19.10*	74	9	**0.19** (0.07-0.32)	43.61*	82
AA	4	**0.22** (0.00-0.11)	3.35	10	10	**0.36** (0.20-0.52)	32.54*	72	2	**0.25** (-0.40-0.90)	10.10*	90
ECO	6	**0.21** (0.00-0.11)	5.16	23	3	**0.41** (0.04-0.78)	16.11	88	6	**0.25** (0.03-0.47)	49.35*	90

Figure 2. *Effect size and 95% confidence intervals for each of the seven outcome variables for studies coded as developer led, involved or indep*
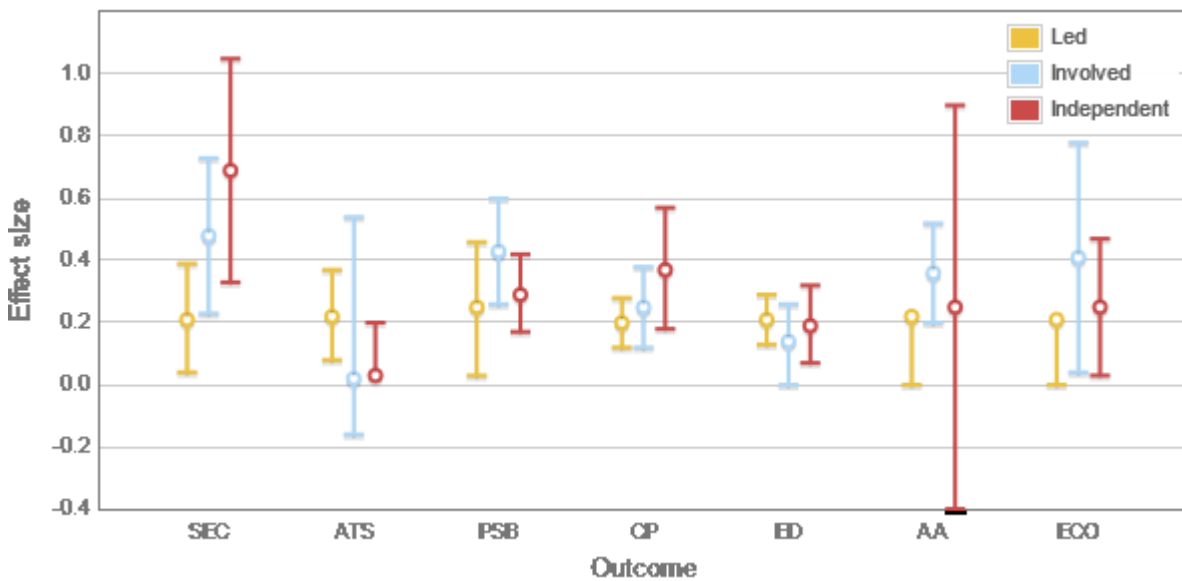
endent.

1

2   Table 7. *Crosstab of developer involvement vs. implementation issues*

Developer involvement	Issues reported- Yes	Issues reported– No	Total
Led	3	20	23
Involved	4	14	18
independent	3	17	20
Total	10	51	61

3

4

5   Table 8. *Mean effects, confidence intervals fit statistics for Cross-cultural transferability*

Outcome	Home				Away			
	n	Effect (95% CI)	Q	I² (%)	n	Effect (95% CI)	Q	I² (%)
SEC	19	**0.56** 0.36-0.77	468.62 *	96	5	**0.06** 0.17-0.28	13.48*	70
ATS	9	**0.20** (0.08-0.32)	10.63	34	1	**0.09** (-0.02-0.20)	-	-
PSB	29	**0.39** 0.27-0.51	30.02*	67	11	**0.20** 0.10-0.30	332.23*	92
CP	30	**0.18** 0.13-0.24	87.15*	68	11	**0.49** 0.20-0.77	297.07*	97
ED	26	**0.21** 0.13-0.29	86.87*	72	7	**0.12** 0.05-0.20	11.02	46
AA	14	**0.28** 0.17-0.40	41.51*	69	2	**0.73** 0.32-1.31	0.39	-
ECO	11	**0.24** 0.13-0.35	22.40*	55	4	**0.31** -0.01-0.62	49.34*	94

9   Figure 3. *Effect size and 95% confidence intervals for each of the seven outcome variables for studies*
10   *coded as home or away.*