

Evaluation of a School-wide Preventive Intervention for Adolescents: The Secondary Social and Emotional Aspects of Learning (SEAL) Programme

Michael Wigelsworth · Neil Humphrey ·
Ann Lendrum

Published online: 5 July 2012
© Springer Science + Business Media, LLC 2012

Abstract Adolescence is a time during which youth are at an increased risk of developing mental health disorders, yet there are relatively few empirically validated preventive interventions designed for implementation in the secondary/high school phase of education. The aim of the current study was to evaluate the impact of one such programme, entitled ‘social and emotional aspects of learning (SEAL)’, over a 2-year period using a quasi-experimental pre-test–post-test control group design. Our sample consisted of 4,443 students (aged 11–12 at the start of the study) attending 41 secondary schools across England—2,442 at 22 SEAL schools and 2,001 at 19 matched control schools. Our outcome measures were the emotional symptoms and conduct problems subscales of the self-report version of the Strengths and Difficulties Questionnaire. Hierarchical linear modelling indicated that the SEAL programme had no discernible impact upon either of these domains of mental health. Further analysis using data from a subset of youth whose baseline scores placed them ‘at-risk’ demonstrated a reduction in difficulties in those attending SEAL schools, but this was matched by a similar trend in those attending control schools. Finally, although there was evidence that student outcomes for conduct problems were mediated by implementation quality, the associated effect size was very small. Variability in student outcomes for emotional symptoms appeared to be unrelated to implementation quality. The implications of these findings for school-based preventive interventions are discussed using three explanatory frameworks—theory failure, implementation failure and research failure.

Keywords Social and emotional learning · School-based prevention · Universal · Evaluation

Introduction

Experiencing social and emotional wellbeing during childhood and adolescence is an important outcome in and of itself (Denham & Brown, 2010; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011) but also has implications for public health because of its associations with academic achievement (Colman et al., 2009), employment (Farrington, Healey, & Knapp, 2004), family and relationship stability (Colman et al., 2009) and other crucial outcomes later in life. Research indicates a rise in child mental health difficulties in the last several decades (Maughan, Iervolino, & Collishaw, 2005). Current estimates suggest that around 1 in 10 children and young people experience clinically significant internalising and/or externalising problems, with higher rates of disorder amongst adolescents compared to children (Green, McGinnity, Meltzer, Ford, & Goodman, 2005). This trend is hypothesised to relate to a range of developmental and educational changes that take place around the beginning of adolescence (McLaughlin & Clarke, 2010). The prevention of such difficulties has become a policy priority, not least because of the significant economic implications. For example, the annual costs associated with mental health disorders in young people are estimated to be \$247 billion in the United States alone (O’Connell, Boat, & Warner, 2009); in England, the cost *per child* of mental health services for complex difficulties is estimated to be £50,000 (Clark, O’Malley, Woodham, Barrett, & Byford, 2005).

As one of the most effective agencies for the promotion of health (including mental health) (Weare, 2010), schools have become the main focus of efforts to reverse the trends

M. Wigelsworth · N. Humphrey (✉) · A. Lendrum
School of Education, University of Manchester, Oxford Road,
Manchester M13 9PL, UK
e-mail: neil.humphrey@manchester.ac.uk

outlined above. As Greenberg (2010) states, “By virtue of their central role in lives of children and families and their broad reach, schools are the primary setting in which many initial concerns arise and can be effectively remediated” (p. 28). To this end, the last several decades have seen an exponential growth in popularity of universal, preventive social and emotional learning (SEL) interventions that are delivered to all children based upon the idiom, ‘an ounce of prevention is worth a pound of cure’. By putting in place mental health provision for all children and young people, it is argued that we can effectively ‘immunise’ them from later difficulties (Merrell & Gueldner, 2010). In theory, such a system is also more cost-effective to implement, since it avoids the costly screening procedures needed to identify those ‘at-risk’ (which, of course, may miss some children in need of targeted support) and the use of highly trained professionals that are often required to deliver targeted interventions (McLaughlin, 2011). As a result, universal preventive approaches are considered to be more sustainable. Also, because universal approaches by definition include all children, their potential for stigmatising participants is reduced (Greenberg, 2010).

In parallel to the growth in popularity of universal SEL interventions, researchers have accumulated a substantial evidence base demonstrating the impact of such programmes on a range of outcomes (including social and emotional competence, mental health difficulties, school attitudes and academic performance) (Durlak et al., 2011; Wilson & Lipsey, 2007). However, not all studies have yielded positive effects; some recent major trials have reported null findings (e.g. Social and Character Development Research Consortium, 2010; Sheffield et al., 2006), and indeed, even well-validated programmes can fail to show main effects of intervention when implemented outside of tightly controlled efficacy trials (e.g. Kam, Greenberg, & Walls, 2003). Amongst the explanations for such variation in outcomes is the issue of implementation quality. Implementation refers to the enactment of an instructional regime (Raudenbush, 2008) and is typically assessed in terms of constructs such as fidelity and dosage (Durlak & DuPre, 2008; Domitrovich, Gest, Jones, Gill, & Sanford-DeRousie, 2010). Examination of these factors has revealed them to be related to variability in outcomes (Durlak & DuPre, 2008), and one-fifth of studies in this field have reported significant difficulties in implementation (Durlak et al., 2011), suggesting that due consideration in evaluation studies is crucial.

In spite of the accumulation of studies noted above, inspection of the evidence base in the field reveals several important gaps. Firstly, there is comparatively little evidence of the effectiveness of interventions in the secondary phase of education (e.g. only 13 % of studies in a recent meta-analysis by Durlak and colleagues (*ibid*) focused on

high school settings). This is of crucial importance given the increased prevalence of mental health disorders in adolescence compared to childhood and the substantial differences between primary/elementary and secondary/high schools in terms of increased size, greater emphasis on ability and competition, and reduced quality of relationships with teachers in the latter (Wigelsworth, Humphrey, & Lendrum, 2012).

Secondly, the majority of the literature reports on efficacy trials carried out under tightly controlled conditions in which schools access high levels of technical support and assistance not normally available to them—meaning that the external validity of many interventions has not been established. This is important given the acknowledged difficulties of bring universal interventions ‘to scale’ (Elias, Zins, Graczyk, & Weissberg, 2003).

A third gap in the SEL evidence base is the relative lack of studies outside of the United States (US). The US have inarguably been the pathfinders in this area, with 83 % of studies in Durlak et al.’s (2011) recent meta-analysis originating there. However, ‘transferability’ cannot be assumed (Blank et al., 2010). The increasing publication of evaluations of SEL interventions in Australia (e.g. Dix, Slee, Lawson, & Keeves, 2012), Germany (e.g. Schick & Cierpka, 2005), and elsewhere (see Humphrey, *in press*, for a review) recently is therefore welcome, but further research that field-tests cultural adaptations of existing programmes and/or home-grown interventions is a priority. This is particularly crucial in the United Kingdom (UK) (the current authors’ location), where a recent systematic review of universal SEL interventions in secondary school settings identified only 3 UK-based studies, each of which was methodologically flawed (Blank et al., 2010); at the same time, a recent international survey ranked the UK bottom of 21 developed countries in relation to child and adolescent wellbeing (UNICEF, 2007).

A further gap in the evidence base is the failure of many evaluations to properly assess the preventive properties of the interventions under scrutiny. Greenberg (2010) cautions that standard analyses, in which effect sizes are computed across an entire sample, may be a poor metric for universal SEL interventions because most participants begin without symptoms, making it unlikely that much change will occur. There are, therefore, additional forms of analysis that may be more appropriate including those in which those youth considered ‘at-risk’ at the beginning of the study (because, for example, they scored at or above the threshold for clinically significant difficulties at baseline) are considered separately, or where odds-ratios are calculated to determine the probability of those attending schools involved in preventive interventions moving from the normal to clinical range on a given scale when compared to those in comparison schools; these are arguably

more robust tests of whether a universal, preventive SEL intervention has achieved its primary aims.

The current study was designed to address the issues and gaps outlined above, in addition to building upon several recommendations made by leading authors in the field, chief amongst which were the need for research to take into account the clustered and hierarchical nature of school-based data (Social and Character Development Research Consortium, 2010), provide data on implementation variability (Durlak & DuPre, 2008), report effects for subgroups of interest (e.g. youth considered to be ‘at-risk’) (Greenberg, 2010) and make greater use of robust outcome measures whose reliability and validity have been proven (Durlak et al. 2011).

The Secondary Social and Emotional Aspects of Learning (SEAL) Programme

SEAL is, “a comprehensive, whole-school approach to promoting the social and emotional skills that underpin effective learning, positive behaviour, regular attendance, staff effectiveness and the emotional health and well-being of all who learn and work in schools” (Department for Children, Schools and Families, 2007, p. 4). It is based on the theoretical framework of emotional intelligence (EI) proposed by Goleman (1996). Goleman’s model of EI is based around five inter- and intra-personal competencies (self-awareness, managing feelings, motivation, empathy and social skills), the promotion of which is hypothesised to a range of favourable outcomes, including improved mental health (DCSF, 2007). Expected improvements in mental health are consistent with an EI model; research has, for example, demonstrated the role played by EI in moderating the effects of chronic stressors on internalising and externalising symptoms in adolescents (Davis & Humphrey, 2012). Such effects are also consistent with the broader logic model for universal SEL interventions (Zins, Bloodworth, Weissberg, & Walberg, 2004), and indeed, published evaluations have borne this out (see Horowitz & Garber, 2006; Wilson & Lipsey, 2007 for meta-analyses pertaining to internalising and externalising symptoms, respectively). However, it is important to note that EI—and in particular the populist version embodied by Goleman’s (1996) work—has also been the subject of considerable criticism. Waterhouse’s (2006) critique summarises the main problems well, including the rather vague definition of EI, a shaky evidence base, overstated claims of the importance of EI, and its promotion as a panacea for social problems.

An overview of the SEAL programme can be found in the “Method” section of this article. It was launched in English secondary schools in 2007 following a brief pilot, the evaluations of which were reported by Smith,

O’Donnell, Easton, & Rudd (2007) and the Office for Standards in Education (OFSTED) (2007). Although generally positive about the programme, these small-scale evaluations both focused on process rather than outcomes, and a larger, more robust inquiry was warranted. To this end, the current authors were commissioned by the English government to conduct a national evaluation of secondary SEAL as the programme was being brought to scale.

In this article, we focus on our analysis of the impact of the programme on the internalising and externalising symptoms of the general population of students, in addition to those at-risk. Additionally, we present analyses that assess the extent to which student outcomes vary as a function of implementation quality. Interested readers are referred to our other publications (Humphrey, Lendrum, & Wigelsworth, 2010; Lendrum, Humphrey, & Wigelsworth, under review; Wigelsworth et al., 2011) for qualitative analyses of implementation variability and impact on other outcomes, such as social and emotional competence.

Method

Design

The study utilised a quasi-experimental, pre-test–post-test control group design. Random assignment to treatment and control groups was not possible because school recruitment to the national SEAL roll out occurred prior to the evaluation being commissioned. SEAL schools agreeing to participate in the research were therefore matched to non-SEAL control schools on the basis on several key socio-demographic indicators. Assessment of outcomes was assessed using a pre-test post-test measurement protocol, with the response variables being emotional symptoms and conduct problems. Hierarchical linear modelling (HLM) was used in order to account for school-level variance and to facilitate the inclusion of a number of appropriate control variables.

Sample

The sample consisted of 2,442 pupils from 22 SEAL schools and 2,001 pupils from 19 control schools from across England. A subsample of 9 of the 22 SEAL schools agreed to participate in additional data collection focusing on implementation quality (see below).

All pupils were aged between 11 and 12 years at the start of the study (Year 7—the first year of secondary education in England). Characteristics of the schools and students are shown in Table 1. In terms of school characteristics, note the similarity between (a) SEAL and control schools and (b) both groups of schools and secondary schools across England. Multiple analyses of variance (MANOVA) and associated

effect analyses demonstrated that the two groups of schools did not significantly differ from one another on any of the measured characteristics (all $p > .05$). One sample t tests and associated effect size analysis confirmed the lack of meaningful differences between SEAL/control schools and national averages in all variables except school size—with schools in our sample being somewhat larger than secondary schools in England on average.

In terms of student characteristics, the table also demonstrates the similarity between pupils in (a) SEAL and control schools and (b) both groups of schools and those across England. In relation to the former, chi-squared tests revealed no significant differences in sex and FSM eligibility between pupils in SEAL and control schools, but differences did emerge in relation to both ethnicity and SEN. However, these were very marginal in terms of magnitude (e.g. a 5.4 % difference in the proportion of white British pupils in relation to ethnicity) and are most likely an artefact of the increased sensitivity of our statistical tests associated with such a large sample. In relation to the latter, one sample t tests confirmed

the lack of significant and meaningful differences between pupils in SEAL/control schools and national averages in all variables.

Analysis of baseline data for our outcome measures (see below) revealed that 593 (13 %) and 714 (16 %) of participants scored in the abnormal ranges for emotional symptoms and conduct problems, respectively; these students were identified as our at-risk sub-sample. These figures are very similar to the national norms for adolescents reported by Green et al. (2005), strengthening claims as to the representativeness of our sample.

Materials

The Strengths and Difficulties Questionnaire (SDQ)

The self-report version of the SDQ (Goodman, 1997) provides a broad behavioural screening profile of adolescents' emotional symptoms, conduct problems, hyperactivity/

Table 1 School and student socio-demographic characteristics in the study sample and national trends

	National average	SEAL schools	Control schools
<i>School characteristics</i>			
Size (no. of pupils)	975 ^a	1,079	1,043
Attainment (% 5 A*–C GCSEs inc. maths and English)	50.7 ^b	41	44.9
Attendance (average days missed)	1.49 ^c	1.24	1.6
% FSM	13.4 ^a	17.3	16.2
SEN	7.18 ^d	7.1	7.3
<i>Student characteristics</i>			
Gender (% female)	49	52	52
Ethnicity (%) ^a			
White (white British/Irish/traveller/any other white background)	79.2	82.3	86.2
Mixed (mixed Caribbean/mixed African/mixed Asian/any other mixed background)	4.2	3.4	1.8
Asian (Indian/Pakistani/Bangladeshi/Chinese/any other Asian background)	9.6	7.2	5
Black (black Caribbean/black African/any other black background)	4.9	2.9	1.4
Other	1.4	1.2	1.0
SEN (%) ^a			
No SEN	81.2	83.7	81.3
School action (SA)	11.6	9.9	14
School action plus (SA +)	5.1	4.4	3.5
Statement (SSEN)	2.1	2.0	1.2
FSM eligibility (% not eligible)	86.9	86.5	88.4

SEN special educational needs, FSM free school meals

Students identified as having SEN in England receive provision at three stages: school action (SA where needs are met within adaptations to normal school practice), school action plus (SA + where input from external professionals (such as educational psychologists) is sought), and statement of SEN (SSEN where a statutory assessment of need has been conducted and a legal document is drawn up outlining the support required to meet the student's needs)

The General Certificate for Secondary Education (GCSE) is the standard exit award for students leaving secondary schools in England. School-level attainment is assessed by the proportion of students who receive at least 5 GCSEs (including English and Maths) at Grades A*–C

^a DCSF (2009), ^b DCSF (2010a), ^c DCSF (2010b), ^d DCSF (2007)

inattention, peer problems and prosocial behaviour, in addition to a ‘total difficulties’ score that is the composite of the four difficulties scales but does not include the prosocial behaviour scale. The analyses reported in the current study focused on the emotional symptoms and conduct problems subscales only, as these were the variables that gave the most directly relevant indices of students’ mental health.

The SDQ consists of a series of statements (e.g. ‘I get very angry and often lose my temper’) to which the respondent indicates a level of agreement on a three-point rating scale (‘Not true’, ‘Somewhat true’, ‘Certainly true’). The self-report version used in this study comprises of 25 items, of which there are five each pertaining to emotional symptoms and conduct problems. The SDQ is amongst the most widely used tools in the field of child and adolescent mental health (Johnston & Gowers, 2005). It is used in many countries including Australia (Hawes & Dadds, 2004) and the United States (Bourdon, Goodman, Rae, Simpson, & Koretz, 2005). SDQ scores demonstrate strong psychometric properties, including factorial validity (established through factor analysis), predictive validity (established through correspondence with independent diagnoses of psychiatric disorders) and internal consistency (average Cronbach’s alpha coefficient is 0.73) (see Goodman, 2001, for a review). In the current study, the Cronbach’s alpha coefficient was 0.69 for the emotional symptoms scale and 0.63 for the conduct problems subscale.

Each subscale is scored from 0 to 10, with higher scores being indicative of greater difficulties. In the self-report version, scores above seven for emotional symptoms and five for conduct problems are considered to abnormal and possible evidence of a clinically recognisable disorder. These thresholds have been validated by research demonstrating a substantially raised probability of independently diagnosed psychiatric disorders when they exceeded (mean odds-ratio: 15.2) (Goodman, 2001).

Assessment of Implementation Variability

As noted above, a subsample of 9 SEAL schools agreed to participate in additional data collection focusing on implementation quality. This aspect of the evaluation presented some unique challenges. Traditional approaches to the assessment of implementation could not be applied because of the open-ended, flexible nature of the intervention (see below). So, for example, it was impossible to assess ‘fidelity’ because there is no single agreed model of implementation. Even though schools were expected to engage in the four key practices noted earlier, the extent to which they did so and the form that this would take was not prescriptively specified. A qualitative approach was therefore utilised, with longitudinal case studies of each school conducted over a 2-year period, involving

interviews with staff, focus groups with students, a variety of observations and document analyses. The rich, detailed data this yielded was analysed using traditional qualitative techniques and is reported elsewhere (Humphrey, Lendrum, & Wigelsworth, 2010; Lendrum, Humphrey, & Wigelsworth, under review). However, we were also able to use the data to form summative judgements about the progress in implementation of each school relative to one another (see Humphrey, Lendrum, & Wigelsworth, 2010). This led to division of the school sample into three implementation clusters: higher quality ($N = 4$), moderate quality ($N = 3$) and lower quality ($N = 2$). Higher quality schools were those where our data suggested that good progress had been made and there had been a comprehensive approach to implementation of the SEAL programme (for example, evidence of activity in the areas outlined under ‘Intervention’ below). Moderate quality schools were those where progress in implementation was mixed, with successes in some but not all aspects of delivery (for example, evidence of consistent delivery of the curriculum element but little work on staff development). Finally, lower quality schools were those where our data suggested little evidence of any real progress in implementation (for example, evidence of a very limited or superficial approach to implementation, applied with little consistency or conviction). These clusters were used as explanatory variables against which to model variation in student outcomes.

Procedure

Survey data from all students in the target cohort in participating schools were collected at the beginning of 2008 (pre-test), and at the beginning of 2010 (post-test). Data were collected in 2009 but were used solely for interim reporting. For each wave of data collection, participating schools were sent a pack of student surveys with administration instructions. A member of staff took responsibility for coordinating the completion of questionnaires (in SEAL schools, this was the designated programme coordinator; in control schools, it was typically the head of year or pastoral care coordinator). Administration of the questionnaires took place in either whole-year (e.g. assemblies) or whole-class (e.g. tutor groups) settings. Any students who had difficulties in completing the surveys were able to solicit support from school staff. Tracking of individual responses over time was achieved through the use of personalised labels on surveys that included students’ names and a unique numerical identifier (this information was used solely for accurate matching and was destroyed once this had been achieved). Once complete, the surveys were collected by courier and delivered to an

independent company who scored and input the data into an electronic database ready for analysis.

Intervention

Schools in the intervention group implemented SEAL between the pre- and post-test dates noted above. Full details on the programme, including guidance materials, can be freely accessed at <http://tinyurl.com/d4selgq>. SEAL may be classified as a ‘multi-component’ programme as the materials include whole-school assemblies, class activities and suggestions to involve the wider community. These activities are based on four key practices, specifically:

- Use of a whole-school approach: “thinking holistically, looking at the whole context including organisation, structures, procedures and ethos, not just at individual pupils or at one part of the picture only” (DCSF, 2007, p. 22). Key components of a whole-school approach include activity in relation to policy development, partnership with parents and the community, promoting a positive school culture and environment, and giving students a voice (Department of Health, 2007);
- Direct and explicit teaching of social and emotional skills: the provision of a series of sessions (referred to as ‘focus groups’) in which students, led by an adult facilitator (e.g. teacher, teaching assistant), take part in activities designed to promote social and emotional competence. These were organised into discrete themes, including ‘learning to be together’ (social skills and empathy), ‘learning about me’ (managing feelings) and ‘keep on learning’ (motivation) designed to be taught throughout the school year;
- The use of teaching and learning approaches that promote a safe and supportive classroom learning environment: this includes ensuring that the pedagogical approach being adopted in ordinary lessons is consistent with SEAL principles, such as using teamwork, co-operative learning and group projects as a means of implicitly promoting social skills (DCSF, 2007);
- Staff training and continuing professional development: examples include coaching and mentoring sessions, discrete training in specific areas (e.g. anger management, assertiveness) and provision of a SEAL working party to drive the programme forward within the school (DCSF, 2007).

However, these are provided as guidance only, and schools are encouraged to, “take from it what they wish” (Weare, 2010, p. 10). Consequently there are, strictly speaking, no ‘critical components’ (Century, Rudnick, & Freeman, 2010) such as might be found in other intervention models. A great deal of variability in implementation can therefore be expected, which puts SEAL at odds with the considerable literatures

emphasising the importance of structure and consistency in programme delivery (e.g. Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2004) and fidelity to a single treatment model (e.g. Carroll et al., 2007; Durlak & DuPre, 2008). It also makes assessment of fundamental aspects of implementation (e.g. fidelity) rather challenging. Nonetheless, the ‘bottom-up’ (as opposed to ‘top-down’) approach to implementation was built into the secondary SEAL guidance from the outset in order to promote local ownership and, ultimately, sustainability (Weare, 2010).

Analytic Strategy

The main analysis of outcomes involved HLM using MLWin version 2.20. HLM is an advancement over related techniques such as multiple regression because it acknowledges the hierarchical (students are nested within schools, which themselves are nested within Local Authorities (LAs), akin to school districts) and clustered (scores of students within a given school are correlated) nature of the data (Paterson & Goldstein, 2007). Failing to account for such structures can seriously underestimate the standard error of the regression co-efficient, potentially leading to spurious results (Twisk, 2006).

For each analysis, the response variable is the post-test score. Pre-test score is controlled for as a student level variable. Consistent with previous approaches to HLM (Gutman & Feinstein, 2008; Humphrey, Lendrum, & Wigelsworth, 2010), the statistical model is constructed as a series of stages. First, an unconditional (‘empty’) model was used to establish the amount of unexplained variance attributable to each level (e.g. LA, school or student). Second, a partial (‘background’) model was used to examine the contribution of school and student characteristics to the variance established in the unconditional model. Third, a final (‘conditional’) model is used to examine the key variable of interest, specifically changes in the response variable as a result of attending a school implementing SEAL.

Results

In accordance with recommendations from a number of reviews (Roth, 1994; Wilkinson, 1999), the dataset was subject to screening prior to inferential testing being conducted, summarised below:

Missing Data

The optimal sample for the study was $N = 4,617$. Of these, 174 cases had missing data that exceeded the acceptable

limits for the SDQ as defined by the developer (greater than three missing responses per scale). For the remaining 4,442, missing items were imputed. Tabulated pattern analysis showed that less than 1 % of missing cases could be attributed to any of the socio-demographic factors included in the analysis, indicating no discernible pattern to missing data. Given this, and the fact that the number of missing cases was very small, case deletion was considered preferable to multiple imputation (Shafer, 1999). Furthermore, estimation techniques serve to support appropriate statistical power to a study; as noted below, the study was more than adequately powered from the outset.

Sample Size

Sample size calculations indicated that a minimum of 16 schools with an average of 100 pupils per school would be required to detect a small effect size ($f^2 = 0.02$) with 14 explanatory variables (the number included in the inferential analysis). The final sample far exceeded this minimum threshold.

Data Requirements and Assumptions

Data were screened to ensure conformity to the assumptions of the inferential analysis conducted (hierarchical linear modelling). Satisfactory results were produced in regards to linearity, normal distribution and independence of errors, homoscedasticity, multi-collinearity and non-zero variance of predictors (Field, 2009; Menard, 1995; Myers, 1990).

Descriptive Statistics

Means and standard deviations of the outcome measures at pre- and post-test are presented in Table 2.

Mean and standard deviation scores of the SEAL and control compared favourably at pre-test for both emotional symptoms ($t(4,441) = 0.822, p > .05$, Cohen's $d = 0.02$) and conduct problems ($t(4,441) = 0.436, p > .05$, Cohen's $d = 0.01$). Post-test scores indicated little change in the two outcome measures for either group. Scores for the at-risk sub-sample demonstrate a reduction in symptoms over time, but no differential effect according to type of school attended (e.g. SEAL versus control).

Inferential Statistics

The two hierarchical linear models developed for the main outcome analyses are shown in Tables 3 and 4.

The empty model shown in Table 3 demonstrates a significant, but small amount of variance associated with the LA level ($\text{var}(v_{0k}) = 0.68, p = 0.04$). The school level

of the model was associated with a statistically insignificant amount of variation in students' emotional symptoms score ($\text{var}(l_{0j}) = 0.025, p = .17$). The magnitude of the effect was also very small, accounting for less than 0.5 % of total variance. Therefore, the variance explained by the model is almost exclusively at the student level ($\text{var}(e_{0i}) = 5.323, p < .01$).

The school-level variable of SEAL status (e.g. SEAL versus control) did not significantly impact on changes in students' emotional symptoms ($b_{0j} = -0.103, p > .05$). Its associated coefficient was indicative of a very small effect, with a reduction of just 0.1 on the emotional symptoms subscale as a result of attending a SEAL school. Chi-squared analysis of the $-2 * \log$ likelihood values of the empty and final models indicated that they did not significantly differ ($\chi^2(1, 4,443) = 1.892, p > .05$), lending further support to the notion that SEAL did not impact upon this outcome variable.

The empty model in Table 4 shows a small and not statistically significant proportion of variance attributable to the LA level ($\text{var}(v_{0k}) = 0.051, p > .05$). School-level variance demonstrates a similarly low contribution (1.5 %) to explaining conduct problems ($\text{var}(u_{0j}) = 0.053, p < .05$) but is nonetheless a statistically significant predictor. As with the previous model, our analysis of changes in pupils' conduct problems demonstrated that the variance was almost exclusively restricted to the student level ($\text{var}(e_{0j}) = 3.502, p < .01$) (97.1 % of all variance).

SEAL status at the school level demonstrated no significant or discernible effect on pupils mental health scores ($b_{0j} = -0.047, p > .05$). The coefficient indicates that student conduct problem scores declined by just 0.047 points as a result of attending a SEAL school. Accordingly, chi-squared analysis of the $-2 * \log$ likelihood values of the empty and final models indicated that they did not significantly differ ($\chi^2(1, 4,443) = 0.68, p > .05$). As above, this was indicative of the failure of SEAL to impact upon the outcome variable.

Analysis of data for the sub-sample of at-risk students was conducted using standard multiple regression in view of the greatly reduced sample size and the general lack of school-level variance identified in the main analyses above. As in the main analyses, the response variable was the post-test score, with pre-test score, SEAL status (e.g. SEAL versus control) and other possible covariates (e.g. sex, SEN status, FSM) controlled for as explanatory variables. These analyses demonstrated that attendance at a SEAL school had no statistically preventive effect for at-risk students for either emotional symptoms or conduct problems (both $p > .05$; see Tables 5, 6).

Further analysis designed to tap the possible preventive effects of SEAL involved calculation of odds-ratios based around the proportion of students moving from the normal

Table 2 Descriptive statistics for participants in the current study

	Pre-test				Post-test			
	SEAL		Control		SEAL		Control	
	All	At-risk	All	At-risk	All	At-risk	All	At-risk
Emotional symptoms	3.57 (2.38)	7.80 (0.95)	3.51 (2.41)	7.80 (1.04)	3.24 (2.31)	5.16 (2.36)	3.38 (2.35)	5.31 (2.50)
Conduct problems	2.42 (2.02)	5.88 (1.07)	2.39 (2.01)	5.95 (1.26)	2.16 (1.88)	3.67 (1.91)	2.24 (1.91)	3.97 (2.09)

Table 3 Hierarchical linear model for the emotional symptoms response variable

Empty model [$\beta_{0ijk} = 3.279 (0.069)$]				Background model [$\beta_{0ijk} = 1.913 (0.553)$]				Full model [$\beta_{0ijk} = 1.957 (0.541)$]			
	Co-efficient β	SE	p		Co-efficient β	SE	p		Co-efficient β	SE	p
<i>LA level</i>	0.068 1.3 %	0.037	0.04	<i>LA level</i>	0.040 1.02 %	0.018	0.02	<i>LA level</i>	0.036 0.90 %	0.017	0.02
				LA attainment	0.016	0.009		LA attainment	0.015	0.009	
<i>School level</i>	0.025 0.47 %	0.026	0.17	<i>School level</i>	0.000 0.00 %	0.000	–	<i>School level</i>	0.000 0.00 %	0.000	–
				FSM eligibility	–0.007	0.006	0.12	FSM eligibility	–0.006	0.006	0.31
				Aggregate SEN	–0.003	0.004	0.22	Aggregate SEN	–0.003	0.004	0.22
				Attainment	–0.003	0.004	0.22	Attainment	–0.003	0.004	0.22
				Unauthorised absence	0.018	0.037	0.31	Unauthorised absence	0.013	0.037	0.31
				Size	0.000	0.000	–	Size	0.000	0.000	–
								SEAL	–0.103	0.74	0.44
<i>Student level</i>	5.323 98.23	0.113	0.00	<i>Student level</i>	3.895 98.98 %	0.083	0.00	<i>Student level</i>	3.894 99.1 %	0.083	0.00
				Gender (if male)	–0.907	0.063	0.00	Gender (if male)	–0.904	0.063	0.00
				Ethnicity (if $p < 0.05$)	Pakistani –0.464	0.192	0.01	Ethnicity (if $p < 0.05$)	Pakistani –0.468	0.192	0.01
					Other. Asian 0.711	0.377 1.977	0.03 0.00		Other. Asian 0.721	0.313 1.983	0.01 0.01
					Traveller 6.443				Traveller 6.475		
				FSM (if eligible)	0.074	0.094	0.22	FSM (if eligible)	0.073	0.094	0.22
				School action SEN	–0.049	0.098	0.30	School action SEN	–0.053	0.098	0.29
				School action plus SEN	0.326	0.156	0.02	School action plus SEN	0.333	0.156	0.02
				Statement SEN	0.385	0.236	0.05	Statement SEN	0.394	0.236	0.04
				Emotional symptoms score pre-test	0.433	0.013	0.00	Emotional symptoms score pre-test	0.433	0.013	0.00
–2 * log likelihood = 20,075.748				–2 * log likelihood = 18,674.041				–2 * log likelihood = 18,672.149			
$\chi^2 (1,401.707, n = 4,443) = , p < 0.001$				$\chi^2 (1.892, n = 4,443) = , p = 0.17$							

to abnormal ranges (as opposed to staying in the normal range) from pre-test to post-test; essentially, this enabled us to examine whether youth attending SEAL schools were

any less likely to develop clinically significant problems during the course of the study than those attending comparison schools. For conduct problems, the comparative

Table 4 Hierarchical linear model for the conduct problems response variable

Empty model [$\beta_{0ijk} = 2.268 (0.066)$]				Background model [$\beta_{0ijk} = 1.428 (0.324)$]				Full model [$\beta_{0ijk} = 1.438 (0.316)$]			
	Co-efficient β	SE	p		Co-efficient β	SE	p		Co-efficient β	SE	p
<i>LA level</i>	0.051 1.4 %	0.035	0.08	<i>LA level</i>	0.00 0 %	0.000	–	<i>LA level</i>	0.000 0 %	0.00	–
				LA attainment	0.004	0.005	0.21	LA attainment			
<i>School level</i>	0.053 1.5 %	0.030	0.04	<i>School level</i>	0.004 0.15 %	0.006	0.26	<i>School level</i>	0.003 0.12 %	0.006	0.31
				FSM eligibility	0.00	0.004	–	FSM eligibility	0.00	0.004	–
				Aggregate SEN	0.00	0.003	–	Aggregate SEN	0.00	0.03	–
				Attainment	–0.002	0.002	0.16	Attainment	0.002	0.002	0.16
				Unauthorised absence	0.008	0.019	0.34	Unauthorised absence	0.006	0.019	0.38
				Size	0.000	0.000	–	Size	0.000	0.000	–
								SEAL	–0.047	0.055	0.20
<i>Student level</i>	3.502 97.1 %	0.075	0.00	<i>Student level</i>	2.604 99.85 %	0.055	0.00	<i>Student level</i>	2.604 99.88 %	0.055	0.00
				Gender (if male)	0.301	0.051	0.00	Gender (if male)	0.301	0.051	0.00
				Ethnicity (if $p < 0.05$)	Bangladeshi –0.945	0.312	0.00	Ethnicity (if $p < 0.05$)	Bangladeshi –0.949	0.311	0.00
					Pakistani –0.346	0.154	0.01		Pakistani –0.345	0.153	0.01
					Indian –0.564	0.244	0.01		Indian 0.560	0.244	0.01
				FSM (if eligible)	0.202	0.077	0.00	FSM (if eligible)	0.217	0.077	0.00
				School action SEN	0.202	0.080	0.01	School action SEN	0.199	0.080	0.01
				School action plus SEN	0.551	0.127	0.00	School action plus SEN	0.556	0.127	0.00
				Statement SEN	–0.008	0.192	0.48	Statement SEN	–0.004	0.192	0.49
				Conduct problems score pre-test	0.440	0.013	0.00	Conduct problems score pre-test	0.441	0.013	0.00
–2 * log likelihood = 18,232.220				–2 * log likelihood = 16,866.276				–2 * log likelihood = 16,865.596			
$\chi^2 (1,365.944, n = 4,443) = , p < 0.001$				$\chi^2 (1, n = 4,443) = 0.68, p = 0.41$							

odds-ratio was 0.991, meaning that students in SEAL schools were marginally *more* likely to move from the normal to the abnormal range than those in comparison schools; for emotional symptoms, the comparative odds-ratio was 1.223, meaning that students in SEAL schools were marginally *less* likely to move from the normal to the abnormal range than those in comparison schools.

Student outcome data for the 9 SEAL schools participating in the additional data collection focusing on implementation were extracted for further analysis. Descriptive statistics can be found in Table 7; inspection of the means appears to demonstrate a possible ‘implementation effect’

for conduct problems but not for emotional symptoms. The data were examined using factorial analysis of variance, which tested the relationship between implementation quality (high, moderate, low) and changes in student outcomes over time (pre-test to post-test). The ANOVAs revealed a significant interaction between implementation quality and time for conduct problems [$F(2,821) = 4.179, p < .05$]. However, the overall effect size observed for this interaction was low (partial $\eta^2 = 0.01$), and post hoc Tukey’s Honestly Significant Differences comparisons failed to distinguish between the three groups (all $p > .05$). There was no interaction between implementation

Table 5 Multiple regression for emotional symptoms in the at-risk sample

	<i>B</i>	SE <i>B</i>	β
SEAL status (if SEAL)	−0.140	0.193	−0.029
Sex (if male)	−1.024	0.203	−0.206**
SEN (if SEN)	−0.041	0.136	−0.012
FSM (if FSM)	0.072	0.267	0.011
Emotional symptoms score pre-test	0.458	0.098	0.187**

$R^2 = 0.085$ ($F(5, 592) = 10.918$, $p < .001$)

** $p < .001$

Table 6 Multiple regression for conduct problems in the at-risk sample

	<i>B</i>	SE <i>B</i>	β
SEAL status (if SEAL)	−0.253	0.148	−0.063
Sex (if male)	0.293	0.155	0.070
SEN (if SEN)	0.166	0.098	0.063
FSM (if FSM)	0.159	0.185	0.032
Conduct problems score pre-test	0.315	0.063	0.182**

$R^2 = 0.054$ ($F(5, 713) = 8.035$, $p < .001$)

** $p < .001$

Table 7 Descriptive statistics for students in schools rated as high, moderate and low in overall implementation quality

	Low	Moderate	High
Emotional symptoms (Pre–post-test difference)	−0.39 (2.10)	−0.41 (2.41)	−0.37 (2.5)
Conduct problems (Pre–post-test difference)	0.17 (1.66)	0.03 (2.41)	−0.34 (1.86)

quality and changes in emotional symptoms [$F(2,821) = 0.018$, $p > .05$].

Discussion

The current study demonstrated that the SEAL programme in English secondary schools failed to impact significantly on the emotional symptoms and conduct problems of either (a) the student population as a whole or (b) a subsample of those deemed to be at-risk by virtue of their pre-test scores. Furthermore, odds-ratio calculations demonstrated that students in SEAL schools were no less likely to develop clinically significant problems over the course of the study than those attending control schools. Although there was evidence that student outcomes for conduct problems were mediated by implementation quality, the associated effect

size was very small. Variability in student outcomes for emotional symptoms appeared to be unrelated to implementation quality. Overall, these findings buck the general trend in the literature, which has hitherto provided relatively consistent evidence of positive outcomes for students participating in universal SEL interventions (Durlak et al., 2011; Horowitz & Garber, 2006; Wilson & Lipsey, 2007). Taken together, the findings of this study make an important contribution to the field. ‘Null’ studies such as this one provide critical information that can inform prevention programme design, delivery and evaluation (Humphrey, in press). Put simply, knowing ‘what doesn’t work’ can be just as vital as knowing ‘what works’. In this final section, we explore the implications of the study, using an adaptation of Raudenbush’s (2008) framework for understanding null results to organise our ideas.

Raudenbush (ibid) suggests two possible explanations for null results when an instructional regime is evaluated—theory failure and implementation failure. To this, we tentatively add a third—research failure. Theory failure is evident when a programme has been implemented as designed and robustly evaluated, but there are problems with the underlying programme theory. SEAL is ostensibly underpinned by EI theory, which as already noted provides a viable basis for school-based prevention efforts, albeit one that has courted some controversy (e.g. Waterhouse, 2006). Perhaps a more likely candidate for theory failure is the actual theory of change underpinning the programme. SEAL was conceived as a loose enabling framework for school improvement rather than a prescriptive, manualised intervention that is more typical of universal SEL in order to promote local ownership and sustainability (Weare, 2010). However, our analyses reported elsewhere (Humphrey et al., 2010; Lendrum, Humphrey, & Wigelsworth, under review) suggested that whilst this more flexible approach was initially welcomed by staff in SEAL schools, ultimately it left them without a clear direction and focus in the implementation process. Indeed, this was illustrated in the current study by the fact that less than half of SEAL schools in our implementation subsample were rated as high quality. Hence, we are minded to recommend that future SEL provision in the UK draws more extensively on the considerable evidence base that speaks to the merits of structure and consistency in programme delivery (Catalano et al., 2004), and fidelity to a core implementation model (Carroll et al., 2007; Durlak & DuPre, 2008) as a means of ensuring positive outcomes. This is not to suggest, however, that such interventions need to be devoid of any scope for adaptation or flexibility; indeed, research suggests that positive outcomes can be achieved with only 60–80 % fidelity for some interventions (Durlak & DuPre, 2008).

The second explanation in Raudenbush’s (2008) repertoire, implementation failure, occurs when a programme

theory is sound and there has been a robust evaluation, but the intervention is not implemented as designed. In the case of SEAL, this issue is tied up with programme theory to a certain extent because of the emphasis on flexibility and local ownership (see above). As a result, it is difficult to determine whether the failure of SEAL to impact on student mental health outcomes is a result of implementation failure, because there is no single core model of implementation against which to assess schools' fidelity, dosage, and so on, using traditional techniques. Our analysis of student outcomes for the 9 SEAL schools involved in the implementation strand of the study indicated that changes in student outcomes for conduct problems were mediated by overall implementation quality, but the associated effect size was very small, and post hoc comparisons failed to distinguish between the low, moderate and high implementation quality groups. Variability in student outcomes for emotional symptoms appeared to be unrelated to implementation quality. In sum, this provides a rather mixed picture. However, our other analyses pertaining to implementation (Humphrey et al., 2010; Lendrum, Humphrey, & Wigelsworth, under review) can provide some useful additional insights. Broadly speaking, we found great variability in implementation of SEAL, with different emphases on various combinations of the four key elements outlined earlier in this article (use of a whole-school approach, direct teaching of social and emotional skills, teaching and learning, and staff development) and varying levels of perceived progress and success. This data also highlighted the increased difficulties in implementing a universal social-emotional learning initiative in a secondary setting—many staff felt that it was not part of their remit/responsibility, and/or that they had to prioritise the academic curriculum because of governmental pressure to increase attainment. Other highlighted issues included the organisational complexity of the secondary school setting, which was seen as creating barriers in terms of consistency of delivery, reinforcement, communication and reduced quality of teacher-pupil relationships (Lendrum, Humphrey & Wigelsworth, under review). To this, we would add the general observation that the more rationalist/technicist ethos of many secondary schools (when compared to primary schools) may act as a further barrier, especially given that EI and related constructs are often perceived as being 'at odds' with rationalist model of schooling.

Ultimately, the issues noted above did not appear to make a difference in outcomes—inspection of residuals charts for our various outcome measures (which plot the deviation of each school's scores from a grand mean) revealed little in the way of school differences and there was no association between different approaches to implementation identified in our fieldwork and other

student outcomes (Humphrey et al., 2010). As noted elsewhere in this paper, the assessment of implementation of SEAL was adapted to the demands of the programme itself; use of traditional techniques would have been futile, largely because there was no single model of implementation against which to assess schools. A key future challenge is for programme developers such as the designers of SEAL to identify what Century, Rudnick, and Freeman (2010) refer to as the 'critical components' of an intervention. This will help implementers distinguish between the 'must dos' and the 'could/should dos' and also enable greater precision in the assessment of implementation.

A third explanation for understanding null results, that of research failure, assumes a sound programme theory and implementation as planned but posits that flaws in the evaluation process influence study outcomes. In this vein, there are several issues worthy of note. Firstly, our study relied solely on student self-report, meaning that we were unable to triangulate our findings against data from other sources (e.g. teachers, parents). However, this is a design featured shared with approximately two-thirds of all universal SEL evaluations (Durlak et al., 2011) and so is unlikely to account for our findings. A second potential problem is the failure to randomly allocate schools to intervention and control groups. It is important to consider the fact that there may have been latent differences between SEAL and control schools in our evaluation—such as motivation, interest in SEL and existing SEL practices—which may have influenced our findings. However, it could be argued that each of these factors could increase, as opposed to decrease, the likelihood of finding differences between schools. Finally, the time frame of our evaluation was limited to 2 years. Whilst this is actually longer than most comparable research in the field (Durlak et al., 2011 reported that 77 % of SEL programme evaluations last less than 1 year), it could be argued that a more complex, multi-component approach such as SEAL could naturally take longer to become fully embedded—and hence influence student outcomes—in participating schools. However, our implementation fieldwork suggested that schools were slowing down (as opposed to ramping up) their activity levels over time, making the likelihood of a 'sleeping effect' emerging post our evaluation rather unlikely.

The above point raises an important issue with regard to the role research plays in the development of educational policy. Several authors (e.g. Torgeson & Torgeson, 2001; Tymms, Merrell, & Coe, 2008) in the UK have called for more rigorous trialling of educational initiatives through randomised controlled trials (RCTs) before they are brought to scale; however, this has met with some resistance, particularly in relation to preventive interventions (Stewart-Brown & Anthony, 2011). To readers in other countries with a stronger focus of 'evidence-based SEL',

such as the US, the notion of bringing an essentially unproven intervention to scale may seem unthinkable. However, in countries like the US, educational policy is largely decentralised, meaning that states and school districts have much more freedom in terms of which educational innovations they choose to adopt. This creates a ‘free market’ situation in which programme developers essentially compete with one another to ‘build a better mousetrap’ (Durlak & DuPre, 2008), and in this context, evidence is king. This contrasts sharply with the UK, where educational policy has largely been centralised, allowing initiatives like SEAL to very quickly become orthodoxies. Consider, for example, the fact that by 2010, when we initially reported our main (null) findings to the government, it was estimated that SEAL was already being implemented in up to 70 % of secondary schools in England (Humphrey, Lendrum, & Wigelsworth, 2010).

As noted at the start of this section, our findings contrast sharply with the dominant trend in the SEL literature. However, it is also important to note the small but significant number of other ‘null’ evaluations that have been reported in the last several years. Amongst the most notable of these are the findings of Social and Character Development Research Consortium’s (2010) multisite RCT of seven different preventive interventions (including several ‘proven’ programmes), and Sheffield et al.’s (2006) RCT exploring different combinations of universal and targeted approaches. Such studies serve as salient reminders that we cannot assume that a programme that has proven to be successful in one context is always going to be successful in another, and the need for what Greenberg (2010) refers to as ‘Type II translational’ research, wherein we move beyond basic questions about efficacy and focus on the factors associated with the successful utilisation of validated interventions (that is, what is it that influences whether programme effects are successfully replicated in typical settings?).

A final point of discussion from the current study stems from the findings in relation to the at-risk subsample. It was evident from our data that the difficulties experienced by at-risk students declined over time, but that was true in both intervention and control schools, indicating that there was no differential preventive effect favouring the SEAL programme. This raises two questions. Firstly, what led to this general decline in difficulties? Secondly, what are the implications for preventive SEL interventions more generally? In relation to the first question, our primary hypothesis is that the decline in difficulties experienced by at-risk youth across the sample was a reflection of their adjustment to secondary school during the period of the study—that is, their initial levels of difficulties were exacerbated by problems adjusting to their new school environment (remembering of course, that the study baseline was taken just a couple of months after the student cohort had entered

secondary education); research would seem to support this notion (McLaughlin & Clarke, 2010). Another strong possibility is that the improvement in students deemed to be at-risk at baseline is simply a result of regression to the mean. In relation to the second question, programme design issues notwithstanding, the relatively ‘light touch’ approach to intervention (in terms of intensity and duration) taken in universal preventive interventions may not be sufficient to impact upon outcomes for those children who are at risk (Greenberg, 2010). A balance between universal (for everyone), targeted (for those considered to be at risk) and indicated (for those already experiencing difficulties) interventions is therefore recommended (Wells, Barlow, & Stewart-Brown, 2003). However, studies that examine the effectiveness of different combinations of these approaches using appropriately rigorous designs are few and far between; indeed, we were only able to find one such example—by Sheffield et al. (2006)—which of course produced null results (see above). Furthermore, if preventive interventions need to be supplemented by targeted/indicated interventions in order to produce desirable effects for at-risk youth, one might also ask (taking a Devil’s advocate position), “what exactly are they preventing?” Clearly, this is a key area for future research to address.

Conclusion

The aim of the current study was to examine the preventive effects of a school-wide SEL intervention in English secondary schools. Our analyses of a nationally representative dataset suggested that the SEAL programme failed to impact significantly on the emotional symptoms and conduct problems of either (a) the student population as a whole or (b) a subsample of those deemed to be at-risk by virtue of their pre-test scores. These findings have important implications in a range of areas, not least the design of universal SEL interventions (including the importance of emphasising structure and consistency in programme delivery), the role of research in policy development (in particular, the need to properly trial educational initiatives before they are brought to scale), and the balance to be struck between universal, targeted and indicated approaches to promoting positive mental health amongst young people.

References

- Blank, L., Baxter, S., Goyder, L., Guillaume, L., Wilkinson, A. S. H., & Chilcott, J. (2010). Promoting wellbeing by changing behaviour: A systematic review and narrative synthesis of the effectiveness of whole secondary school behavioural interventions. *Mental Health Review Journal*, 15, 43–53.

- Bourdon, K. H., Goodman, R., Rae, D. S., Simpson, G., & Koretz, D. S. (2005). The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 557–564.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40–49.
- Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2004). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *The Annals of the American Academy of Political and Social Science*, 591, 98–124.
- Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, 31, 199–218.
- Clark, A. F., O'Malley, A., Woodham, A., Barrett, B., & Byford, S. (2005). Children with complex mental health problems: Needs, costs and predictors over one year. *Child and Adolescent Mental Health*, 10, 170–178.
- Colman, I., Murray, J., Abbott, R. A., Maughan, B., Kuh, D., Croudace, T. J., et al. (2009). Outcomes of conduct problems in adolescence: 40 year follow-up of national cohort. *British Medical Journal*, 338, a2981.
- Davis, S. K., & Humphrey, N. (2012). Emotional intelligence as a moderator of stressor–mental health relations in adolescence: Evidence for specificity. *Personality and Individual Differences*, 52, 100–105.
- Denham, S., & Brown, C. (2010). “Plays nice with others”: Social–emotional learning and academic success. *Early Education & Development*, 21, 652–680.
- Department for Children, Schools and Families. (2007). *Social and emotional aspects of learning in secondary schools: Guidance for schools*. London: DCSF Publications.
- Department for Children, Schools and Families. (2009). *Schools, pupils, and their characteristics, January 2009*. London: DCSF Publications.
- Department for Children, Schools and Families. (2010a). *Key stage 4 attainment by pupil characteristics, in England 2008/9*. London: DCSF Publications.
- Department for Children, Schools and Families. (2010b). *Pupil absence in schools, including pupil characteristics: 2008/9*. London: DCSF Publications.
- Department of Health. (2007). *Introduction to the national healthy schools programme*. London: Department of Health.
- Dix, K. L., Slee, P. T., Lawson, M. J., & Keeves, J. P. (2012). Implementation quality of whole-school mental health promotion and students' academic performance. *Child and Adolescent Mental Health*, 17, 45–51.
- Domitrovich, C., Gest, S. D., Jones, D., Gill, S., & Sanford-DeRousie, R. M. (2010). Implementation quality: Lessons learned in the context of the Head Start REDI trial. *Early Childhood Research Quarterly*, 25, 284–298.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82, 405–432.
- Elias, M. J., Zins, J. E., Graczyk, P., & Weissberg, R. P. (2003). Implementation, sustainability, and scaling up of social–emotional and academic innovations in public schools. *School Psychology Review*, 32, 303–319.
- Farrington, D. P., Healey, A., & Knapp, M. (2004). Adult labour market implications of antisocial behaviour in childhood and adolescence: Findings from a UK longitudinal study. *Applied Economics*, 36, 93–105.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.
- Goleman, D. (1996). *Emotional intelligence: Why it can matter more than IQ*. London: Bloomsbury.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586.
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337–1345.
- Green, H., McGinnity, A., Meltzer, H., Ford, T., & Goodman, R. (2005). *Mental health of children and young people in Great Britain*. Cardiff: Office for National Statistics.
- Greenberg, M. T. (2010). School-based prevention: Current status and future challenges. *Effective Education*, 2, 27–52.
- Gutman, L., & Feinstein, L. (2008). *Children's well-being in primary school: Pupil and school effects*. London: Centre for Research on the Wider Benefits of Learning.
- Hawes, D. J., & Dadds, M. R. (2004). Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Australian and New Zealand Journal of Psychiatry*, 38, 644–651.
- Horowitz, J. L., & Garber, J. (2006). The prevention of depressive symptoms in children and adolescents: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 74, 401–415.
- Humphrey, N. (in press). *Social and emotional learning: A critical appraisal*. London: Sage.
- Humphrey, N., Lendrum, A., & Wigelsworth, M. (2010). *Secondary social and emotional aspects of learning (SEAL): National evaluation*. Nottingham: Department for Education.
- Johnston, C., & Gowers, S. (2005). Routine outcome measurement: A survey of UK Child and Adolescent Mental Health Services. *Child and Adolescent Mental Health*, 10, 133–139.
- Kam, C. M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science*, 4, 55–63.
- Maughan, B., Iervolino, A. C., & Collishaw, S. (2005). Time trends in child and adolescent mental disorders. *Current Opinion in Psychiatry*, 18, 381–385.
- McLaughlin, K. A. (2011). The public health impact of major depression: A call for interdisciplinary prevention efforts. *Prevention Science*, 12, 361–371.
- McLaughlin, C., & Clarke, B. (2010). Relational matters: A review of the impact of school experience on mental health in early adolescence. *Educational and Child Psychology*, 27, 91–103.
- Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks: Sage.
- Merrell, K. W., & Gueldner, B. A. (2010). *Social and emotional learning in the classroom: Promoting mental health and academic success*. London: Guilford Press.
- Myers, R. (1990). *Classical and modern regression with applications*. Boston: Duxbury.
- O'Connell, E., Boat, T., & Warner, K. (2009). *Preventing mental, emotional and behavioral disorders among young people: Progress and possibilities*. Washington, DC: National Academies Press.
- Office for Standards in Education. (2007). *Developing social, emotional and behavioural skills in secondary schools*. London: Office for Standards in Education.
- Paterson, L., & Goldstein, H. (2007). New statistical methods for analysing social structures: An introduction to multilevel models. *British Educational Research Journal*, 17, 387–393.

- Raudenbush, S. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal*, 45, 206–230.
- Roth, P. (1994). Missing data: A conceptual review for psychologists. *Personnel Psychology*, 47, 537–560.
- Schick, A., & Cierpka, M. (2005). Faustlos: Evaluation of a curriculum to prevent violence in elementary schools. *Applied and Preventive Psychology*, 11, 157–165.
- Shafer, J. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, 8, 3–15.
- Sheffield, J. K., Spence, S. H., Rapee, R. M., Kowalenko, N. K., Wignall, A., Davis, A., et al. (2006). Evaluation of universal, indicated and combined cognitive-behavioural approaches to the prevention of depression among adolescents. *Journal of Consulting and Clinical Psychology*, 74, 66–79.
- Smith, P., O'Donnell, L., Easton, C., & Rudd, P. (2007). *Secondary social, emotional and behavioural skills pilot evaluation*. Nottingham: DCSF Publications.
- Social and Character Development Research Consortium. (2010). *Efficacy of school-wide programs to promote social and character development and reduce problem behavior in elementary school children*. Washington, DC.
- Stewart-Brown, S., & Anthony, R. (2011). Should randomised controlled trials be the “gold standard” for research on preventive interventions for children? *Journal of Children's Services*, 6, 228–235.
- Torgeson, C., & Torgeson, D. (2001). The need for randomised controlled trials in educational research. *British Journal of Educational Studies*, 49, 316–328.
- Twisk, J. (2006). *Applied multi-level analysis*. Cambridge: Cambridge University Press.
- Tymms, P., Merrell, C., & Coe, R. (2008). Educational policies and randomised controlled trials. *The Psychology of Education Review*, 32, 3–7 & 26–29.
- UNICEF. (2007). *An overview of child well-being in rich countries*. Florence, Italy: UNICEF.
- Waterhouse, L. (2006). Multiple intelligences, the Mozart effect, and emotional intelligence: A critical review. *Educational Psychologist*, 41, 207–225.
- Weare, K. (2010). Mental health and social and emotional learning: Evidence, principles, tensions, balances. *Advances in School Mental Health Promotion*, 3, 5–17.
- Wells, J., Barlow, J., & Stewart-Brown, S. (2003). A systematic review of universal approaches to mental health promotion in schools. *Health Education*, 103, 197–220.
- Wigelsworth, M., Humphrey, N., & Lendrum, A. (2012). A national evaluation of the impact of the secondary social and emotional aspects of learning (SEAL) programme. *Educational Psychology*, 32, 213–238.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594–604.
- Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine*, 33(Suppl), S130–S143.
- Zins, J. E., Bloodworth, M. R., Weissberg, R. P., & Walberg, H. J. (2004). The scientific base linking social and emotional learning to school success. In J. E. Zins, R. P. Weissberg, M. C. Wang, & H. J. Walberg (Eds.), *Building academic success on social and emotional learning: What does the research say?* (pp. 3–22). New York: Teachers College Press.