

Case Study 1: An Evidence-Based Practice Review Report

Theme: School Based Interventions for Learning

How effective is the Zippy's Friends programme at improving social and emotional outcomes for children?

Summary

A systematic literature review was conducted to evaluate the effectiveness of the Zippy's Friends programme at improving social and emotional outcomes for children. Zippy's Friends is a universal school-based intervention, delivered over 24 weekly sessions, targeting students aged 5-8 years. The intervention revolves around stories of three main characters and an imaginary pet insect called Zippy, covering themes of understanding feelings and emotions, communication, relationships, conflict resolution and coping with change and loss. Zippy's Friends aims to provide a range of coping skills so when faced with difficulties children will have the skills to manage and, in turn, psychological outcomes will be improved. Searches in Web of Science, SCOPUS, psychINFO and ERIC identified five relevant studies for review, conducted between 2006 and 2014. The studies were reviewed using an adapted version of Kratochwill's (2003) Coding Protocol and evaluated using Gough's (2007) Weight of Evidence Framework. Four studies were given overall weightings of medium and one of low, indicating a lack of rigorous methodology suitable to answering the review question. Effect sizes ranged from not practically significant to large across coping, social, emotional and behavioural measures. This review evaluates these studies and suggests Zippy's Friends to be a promising intervention for supporting children's social and emotional needs. Implications for future practice and research are offered.

Introduction

Rationale and Relevance to EP Practice

In the current educational climate, schools are required to foster more than just academic outcomes, reflected in the areas of need identified in the code of practice; not only is cognition and learning an area for development, but also social, emotional and mental health, physical and sensory needs, and communication and interaction (DfE, 2014). Greenberg et al. (2003) note the “increased economic and social pressures on families; weakening of community institutions that nurture children’s social, emotional, and moral development” (p. 467), which highlights the increased importance of supporting children’s psychological well-being. They identify the importance of schools as locations for this development, with their scope being far broader than merely supporting learning.

Additionally, governmental guidance (DfES, 2001) emphasises the importance of preventative work and whole school approaches to supporting mental health, rather than merely supporting children who are already displaying symptomology. Additional benefits in terms of academic outcomes that arise from supporting mental health are also acknowledged, giving further grounding for its importance. In regards to the role of educational psychologists, early intervention and preventative work are areas particularly valued by other practitioners (Farrell et al., 2006) and therefore the requirement to make recommendations of interventions that will provide universal support in this area of need is inevitable.

The Zippy’s Friends programme is one such preventative programme. It aims to teach coping skills, with the purpose of improving children’s ability to cope with negative life events. Its ultimate aim is to reduce negative social, emotional and

mental health problems, which could otherwise arise as a result. Clarke, Bunting and Barry (2014) state the programme is currently being used in 27 countries, which highlights the need to evaluate its effectiveness to ensure the increasing popularity is underpinned by a strong evidence base. There is an increased emphasis on evidence-based practice (Cottrell, 2002) and educational psychologists are accountable for their recommendations, such as that of Zippy's Friends.

Zippy's Friends

The Zippy's Friends programme was originally created through joint working between the global suicide prevention agency Befrienders International (now known as Befrienders Worldwide) (<http://www.befrienders.org/>), a group of researchers and a company with expertise in educational resources. The programme was originally called 'Reaching Young Europe', but was renamed as Zippy's Friends when the charity Partnership for Children (<http://www.partnershipforchildren.org.uk/>) became the distributor and application was allowed outside of Europe.

Zippy's Friends is a universal, school-based intervention, which targets children age 5 to 8 years. It is delivered by class teachers and consists of 24 weekly sessions, lasting around 50 minutes. The intervention revolves around stories of three main characters and Zippy (their imaginary pet insect). It consists of six modules, which cover the themes of understanding feelings and emotions, communication, relationships, conflict resolution and coping with change and loss. Each module consists of 4 sessions, which are underpinned by the 'goal' of that module (Appendix A). Repetition is used throughout the programme to reinforce key ideas and reviews of previous sessions are carried out each lesson.

A key objective of the programme is to equip children with a range of coping skills, emphasising the importance of children exploring their own preferences in strategies and identifying solutions themselves, rather than the programme dictating what should be done. Shared support and seeking advice is another focus, opposed to advocating individual coping strategies.

Zippy's Friends was evaluated in 2000 by Mishara and Ystgaard who found that although several social skills were improved, there were no benefits relating to coping skills. Due to this, further revisions to the programme were made to address this flaw. A systematic review of the findings since this time will assess whether these adaptations have led to the desired improvements in coping outcomes, or whether social skills are still the predominant benefit gained from the intervention.

Psychological Basis

Lazarus and Folkman's (1984) theory of coping underpins the Zippy's Friends intervention. This suggests there are two main types of coping: emotion-focussed coping, which involves regulating negative emotional reactions to stressors, and activity-focussed coping, which involves removing or reducing the stressor. Activity-focussed coping is beneficial due to addressing the cause of the concern; however, emotion-focussed coping may be the only possibility if a stressor is out of the person's control. Perception of controllability also has an influence, as individuals may opt for emotion-focussed coping due to a perception of lack of control, which may not reflect reality, as seen in pessimistic individuals (Nes & Segerstrom, 2006).

Coping efficacy is a subjective view of our ability to cope with demands of a stressor and the emotions evoked by it. This relates to beliefs about our ability to act to cause change or cope, rather than believing outcomes will be positive due to external causes or chance (Scheier & Carver, 1987). Lazarus and Folkman's (1984) transactional model of coping posits that ability to cope will be determined by the evaluation of our ability to manage a stressor to avoid negative outcomes that would impede emotional well-being. Coping efficacy is assumed to have a bi-directional relationship with coping strategies and outcomes, as implementation of coping strategies that lead to beneficial outcomes will in turn enhance coping efficacy. In contrast, negative outcomes will decrease coping efficacy and decrease the likelihood of coping strategies being implemented.

This theory of coping explains the differential consequences of negative life events, as it suggests it is not merely due to whether negative events are encountered, but also how these are cognitively appraised and whether an individual believes they have the ability to cope, as this in turn influences the actions taken in order to cope. Zippy's Friends is based upon the premise that children will face negative events at some point during their lives and it is therefore important to equip them with coping strategies as this will offer specific skills, as well as improving coping efficacy beliefs. This is in line with Hampel and Petermann (2006) who found that emotional and behavioural problems were negatively associated with implementation of coping skills.

Segal (1983) highlights that there is no singular coping strategy that will be most beneficial in all situations. Zippy's Friends acknowledges this by identifying a range

of coping strategies and encouraging children to ascertain which strategy to employ themselves.

Review Question

This systematic literature review will address the question ‘how effective is the Zippy’s Friends programme at improving social and emotional outcomes for children?’ Whether coping skills are an additional benefit gained from the programme (since its revision), or whether social skills remain the main outcome, will be noted as part of this review.

Literature Search

On 3rd February, a literature search was conducted using the following electronic databases: PsychINFO, ERIC, SCOPUS and Web of Science Core Collection. “Zippy’s Friends” was used as the search term, entered between quotation marks to search for it as a phrase. Searches were refined to the title and abstract or topic, depending on the database search options. The initial search yielded 32 results (Web of Science Core collection: 6, SCOPUS: 11, ERIC: 6 and PsychINFO: 9). 17 papers were removed as duplicates and the remaining were screened against inclusion/exclusion criteria. Initially the title and abstract were screened and the full text was reviewed when clarification was required. Table 1 details the inclusion/exclusion criteria.

Table 1

Inclusion and exclusion criteria

	Inclusion Criteria	Exclusion Criteria	Rationale
1) Publication type	Published journal article or thesis	Literature is not in the form of a published journal or thesis	Published journals and theses will have undergone an ethics procedure, whereas other literature may not have undergone the same level of ethical approval and supervision
2) Intervention	Solely implementing Zippy's Friends	Studies not solely implementing Zippy's Friends	Focus of the review is on Zippy's Friends. If Zippy's Friends was not solely implemented, the reviewer would be unable to distinguish which effects were specifically due to the Zippy's Friends intervention
3) Participants	Children/young people between the ages of 5 and 8 years	Children below 5 years of age or above 8 years of age	Zippy's Friends intervention is aimed at the 5-8 year age range

	Inclusion Criteria	Exclusion Criteria	Rationale
4) Outcomes	<p>Social /emotional/ mental health outcomes. Outcomes are in relation to the children. Outcome measures are quantitative.</p>	<p>Outcomes do not relate to social/ emotional/ mental health outcomes of the children e.g. outcomes concern parents or 'classroom climate'. Outcome measures are qualitative.</p>	<p>The focus of the review is on the social/ emotional/ mental health outcomes for children. Quantitative data allows effect sizes to be calculated and comparisons made across studies.</p>
5) Setting	<p>School-based</p>	<p>Any non-school based setting</p>	<p>Zippy's Friends is a school based intervention</p>
6) Study design	<p>Must be a group design that reports between group outcomes and gives pre and post intervention data</p>	<p>Not a group based design. Study does not have pre and post intervention outcome data.</p>	<p>Pre and post measures and between group outcomes allow the impact of the intervention to be assessed and studies compared. Zippy's Friends is a universal programme that should be implemental at class group level</p>

	Inclusion Criteria	Exclusion Criteria	Rationale
7) Language	Written in English. No restriction on the country in which the research was conducted.	Not written in English.	Cost and time limits do not allow for translation, as the reviewer does not speak other languages. If studies conducted in other countries have been published in English, the accuracy of any translation can be assumed
8) Data	Data reported is not used in another study included in the review	Data reported is included in another study included in the review	If two studies are reporting the same data, it would lead to replication during the review which could lead to a bias in findings

Ten papers were excluded on the basis of the inclusion/exclusion criteria (rationale for exclusion in Appendix B). See Figure 1 for the database search/literature screening process and Table 2 for the list of studies included in the full review.

Figure 1: Database search and literature screening process

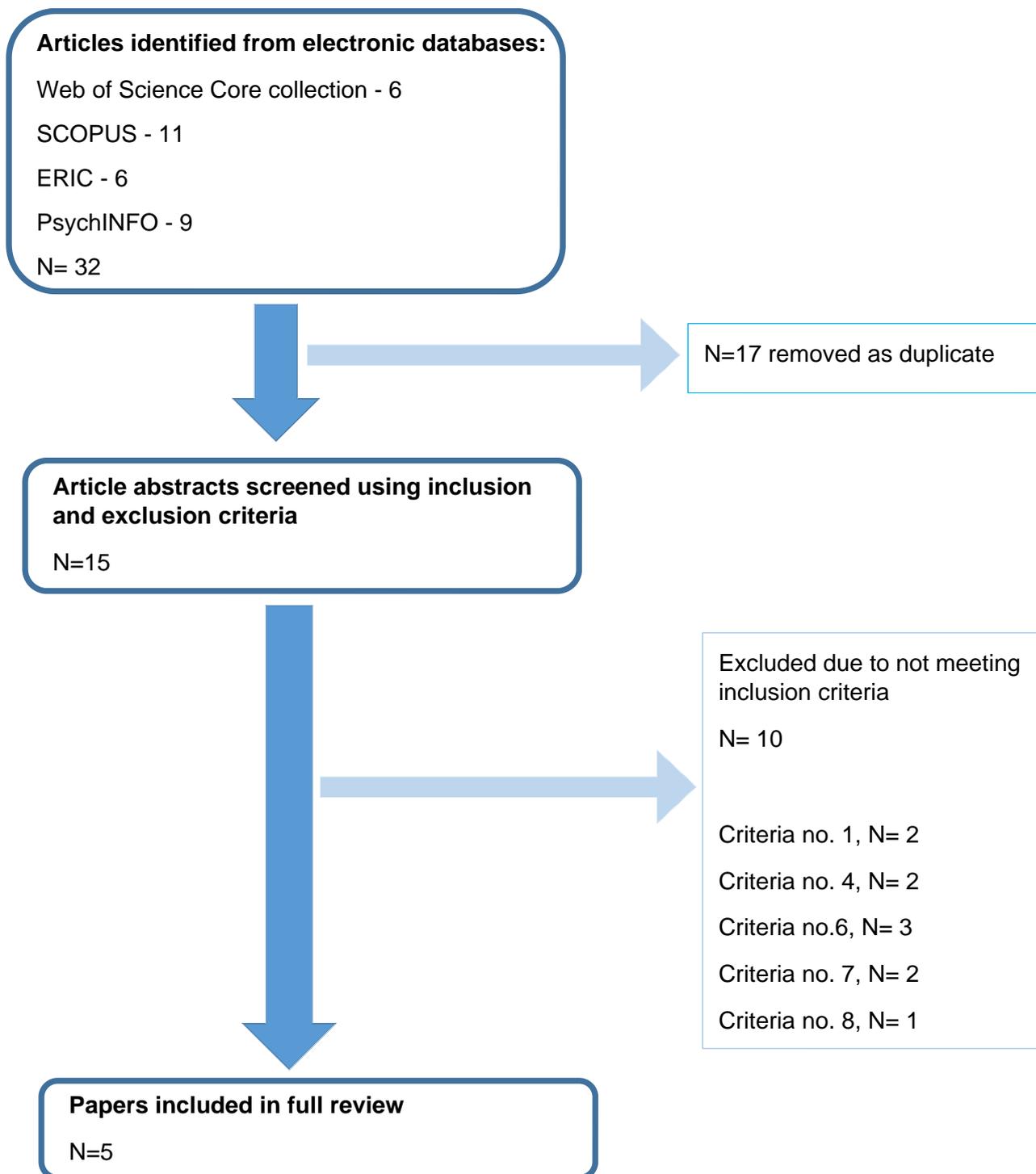


Table 2***Final studies included in the systematic review***

Clarke, A. M., Bunting, B., & Barry, M. M. (2014). Evaluating the Implementation of a School-Based Emotional Well-Being Programme: A Cluster Randomized Controlled Trial of Zippy's Friends for Children in Disadvantaged Primary Schools. *Health Education Research, 29*(5), 786-798.

Dufour, S., Denoncourt, J., & Mishara, B. L. (2011) Improving Children's Adaptation: New Evidence Regarding the Effectiveness of Zippy's Friends, a School Mental Health Promotion Program, *Advances in School Mental Health Promotion, 4*(3), 18-28.

Holen, S., Waaktaar, T., Lervåg, A., & Ystgaard, M. (2012). The effectiveness of a universal school-based programme on coping and mental health: a randomised, controlled study of Zippy's Friends, *Educational Psychology, 32*(5), 657-677.

Mishara, B., & 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*(1), 110-123.

Rodker, J. D. (2013). Promoting Social-Emotional Development of Children During Kindergarten: A Zippy's Friends Program Evaluation. (unpublished doctoral thesis). Pace University, New York, America.

Critical review of the studies

The studies eligible for full review were summarised (Appendix C) and their methodological quality appraised using an adapted version of the Kratochwill (2003) coding protocols from the APA Task Force on Evidence Based Interventions in

School Psychology for group-based designs. The rationale for excluding elements of the coding protocol is detailed in Appendix D and an example completed protocol is available in Appendix E. Gough’s (2007) Weight of Evidence (WoE) framework approach was used (Table 3), which provided criteria to be applied during the review. The framework appraises studies in terms of methodological quality (WoE A), methodological relevance (WoE B), relevance to the review question (WoE C) and overall WoE (WoE D). WoE A evaluates studies against agreed criteria, whereas WoE B and C allow the researcher to consider aspects that are of particular relevance to the review question. WoE D is the average of WoE A, B and C and allows for a clear comparison of the overall weight of each study. Table 4 details the WoE ratings given for the five studies included in this review.

Table 3
The Weight of Evidence framework and areas of focus according to Gough (2007)

Weight of Evidence A	Weight of Evidence B	Weight of Evidence C	Weight of Evidence D
Quality of Methodology: Accuracy, coherency and transparency of evidence in relation to other studies of that type	Relevance of Methodology: appropriateness of methodology and research design for answering the review question	Relevance of evidence to the review question: appropriateness of the focus and details of the study for answering the review question	Overall weight of evidence: the overall average for A, B & C to give the degree to which the study answers the review question

Table 4
Overall weight of evidence

Studies	WoE A	WoE B	WoE C	WoE D
Clarke, Bunting & Barry (2014)	Medium (2)	Medium (2.3)	High (3)	Medium (2.4)
Dufour , Denoncourt & Mishara (2011)	Low (1.2)	Medium (2)	Medium (2)	Medium (1.7)
Holen , Waaktaar , Lervåg & Ystgaard (2012)	Medium (1.8)	Medium (2)	Medium (2)	Medium (1.9)
Mishara & Ystgaard (2006).	Low (1.4)	Medium (1.5)	High (3)	Medium (2)
Rodker (2013)	Low (0.4)	Medium (1.6)	Low (1)	Low (1)

Average WoE scores in Table 4 have been rounded to one decimal place.

Appendices F-I detail the rationale for weightings given and provide summary tables showing how WoE were determined.

Participant characteristics

Studies were conducted in a number of countries: Ireland (Clarke et al., 2014), Canada (Dufour et al., 2011), Norway (Holen et al., 2012), America (Rodker, 2013), Denmark and Lithuania (Mishara & Ystgaard, 2006). This is likely to reflect the programme’s development from a programme initially targeting Europe. Consistent findings would indicate a level of generalisability across settings. However, caution must still be taken when advising the use of the intervention outside of these countries.

Participants were all aged 5-8 years, as stipulated in the inclusion criteria, which is the recommended age range for Zippy’s Friends. All studies gave information relating to the gender of child participants. Rodker’s (2013) study had the largest

gender bias, which was towards boys (58.4%). All other studies had slightly more boys, but this gender bias was very small, ranging from 50.7% boys in Holen et al.'s (2012) study to 52.9% boys in Dufour et al.'s (2011) study. This small difference is unlikely to have affected findings.

In regards to the socio economic status (SES) of participants, no information was given by Dufour et al. (2011). Rodker (2013) and Mishara & Ystgaard (2006) stated schools participating were comparable in terms of their SES, but no further information was given. Holen et al. (2012) determined SES as relating to parental education level, which they categorised as 'low' (up to and including high school, which reflected 454 children's parents) or 'high' (higher education, which reflected 818 children's parents). This dichotomy was used due to the low level of social inequality in Norway. This contrasts Clarke et al.'s (2014) study in which all participating schools were given disadvantaged status from the Department of Education and Skills. This may have affected findings, as prior research shows children from low SES families are more likely to have social-emotional difficulties (Ermisch, 2008; Power & Matthews, 1997).

Holen et al. (2012) used the same SES dichotomy based on level of education for teachers and identified 411 as 'low' and 756 as 'high'. Dufour et al. (2011) was the only other study to give information pertaining to the teachers; 35 teachers participated, with 29 having had more than eight years teaching experience and 28 had worked at their current school for at least three years. The experience of the teachers may have improved the accuracy of intervention implementation, as current classroom management and rapport with pupils may already be well established in

classes with experienced teachers. The other studies reviewed did not report demographic information on parents of the children or teachers. Future studies implementing Zippy's Friends would benefit from gathering this additional information, as it may allow researchers to establish underlying reasons for when Zippy's Friends is most effectively implemented.

Sample size

The number of participants varied greatly from 127 (Rodker, 2013) to 1297 (Holen et al., 2012). Although, it should be noted that due to different designs, the total number of participants does not directly relate to the size of intervention and control groups. For example, in three studies participants comprised one intervention and one control group (Dufour, et al., 2011; Holen et al., 2012; Rodker, 2013), whereas children in Clarke et al.'s (2014) study were split between control and two intervention types. Additionally, students in Mishara and Ystgaard's (2006) study were split across two implementations in Denmark and Lithuania, for which the Denmark implementation had control and intervention groups, whereas the Lithuania implementation lacked a control group for some measures.

A power analysis was conducted to determine whether studies had used enough participants. There has not been a meta-analysis specifically on preventative school-based programmes to promote coping skills, but a meta-analysis of school-based universal programmes to support student's social and emotional learning found effect sizes between 0.22 and 0.57 (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). This indicates effect sizes are likely to be small to medium. Cohen's (1992) table was used to determine whether studies had a sufficiently large

'N', using 0.05 as the level of significance and 'tests of mean differences' as the type of analysis (as the primary tests used were between-participant comparisons of control versus intervention groups). Rodker (2013) was close to meeting the suggested N for identifying a medium effect size. Three of the studies had enough participants to detect a medium effect size, but not small (Clarke et al., 2014; Dufour et al., 2011; Mishara & Ystgaard, 2006). Holen et al. (2012) was the only study with enough participants to detect a small effect. Not having a sufficient 'N' limits the likelihood of detecting effects and is reflected in WoE A.

Research Design

Only quantitative studies with pre and post intervention data were included in the review. Rodker (2013) utilised a school that was already implementing Zippy's Friends and therefore assignment to conditions was not random. Dufour et al. (2011) and Mishara and Ystgaard (2006) did not indicate randomisation of allocation to groups. Clarke et al. (2014) and Holen et al. (2012) both randomised allocation of schools to intervention and control groups. Holen et al, (2012) matched schools before randomisation occurred to ensure group equivalence, whereas Clarke et al. (2014) did post-hoc analysis to determine there was no significant difference between groups.

Only Clarke et al. (2014) used both post and follow-up tests. Post-tests were carried out immediately after the intervention had ceased for all reviewed studies except Rodker (2013) who took post-intervention measurements two months after the intervention had finished. This is reflected in WoE B, as pre, post and follow-up measures are the ideal because it allows for assessment of both the immediate and prolonged impact of the intervention. This is useful for practitioners recommending

interventions, as schools will prefer interventions that have a lasting impact. Post-intervention measures being implemented immediately after the intervention are beneficial because outcomes may dissipate with time or there may be other factors influencing the outcome measures during any gap in time between the end of the intervention and post-intervention assessment. Therefore, caution must be taken when interpreting Rodker (2013)'s findings.

Control group

Four of the studies used a 'no intervention' control group (Dufour et al., 2011; Holen et al., 2011; Mishara & Ystgaard, 2006; Rodker, 2013). In Rodker's study it was noted that the control group received social-emotional learning as part of the curriculum, but this was considered a 'no intervention' control group in this review, as no additional intervention/attention placebo was in place above the typical curriculum. Clarke et al. (2014) was the only study to use an active control group and a no intervention control group. The active control group was given the Zippy's Friends materials to be utilised as a resource. However, when reviewing implementation fidelity, the researchers found both were implemented accurately, so both intervention groups' data were analysed together. Having an active comparison group was the most methodologically relevant design for this review, as it allows for comparison of effects against both maturation effects and against an alternative intervention; this has contributed to overall WoE B ratings (Appendix G).

Measures

A range of measures were used across the five studies which can be considered beneficial in terms of capturing the range of benefits the Zippy's Friends intervention may cause. Additionally, it allows the researcher to gauge whether there are coping

benefits additional to anticipated social skills benefits, which was an aim of the last revision of Zippy's Friends. However, it also means the replication across studies is limited, making comparisons less comprehensive. Table 5 details the range of measures used and gives additional information, including reliability statistics. This information has been sought from the reviewed studies and other publications detailing psychometric properties of the measures, where this was not given in the studies reviewed. Behavioural measures were included in this review as their subscales related to social and emotional constructs.

Table 5
Details of measures used

Measure	Details	No. of items	Reliability		Internal consistency	Studies used in
			Test-retest	Inter-rater		
Coping Strategy Measures						
Schoolagers Coping Strategies Inventory (Ryan-Wenger, 1990)	Questionnaire. Child rates from 0-3 how often they do each behaviour when stressed/ worried/ nervous. Child then rates on a 4 point scale how helpful the behaviour is.	26			Dufour et al. (2011) 0.84	Mishara & Ystgaard (2006) Dufour et al. (2011) Used French translation Mishara & Ystgaard (2006) Teachers completed
Observation based on the Schoolagers Coping Strategies Inventory (Ryan-Wenger, 1990)	Teacher or parent to identify the most important conflict or problem the child has recently experienced, describe it and how the child had reacted and rate on the form which coping strategies were used.	26			Mishara & Ystgaard (2006) unknown Dufour et al. 2011) 0.58-0.61	Dufour et al. (2011) Parents completed Holen et al. (2012) Norwegian translation.
Kidcope questionnaire (Spirito, Stark, and Williams, 1988)	Measures 10 coping strategies (distraction, cognitive restructuring, social withdrawal, blaming others, problem solving, wishful thinking, emotional expression, social support, self-criticism, and resignation). Children's response items dichotomous; whether they would use the coping strategy. Parent's had 4 response options: Not at all, sometimes, a lot of the time, almost all the	15 (children's version) 12 (adolescent version)	0.41 to 0.83 over a 3-7 day period. 0.15-0.43 over a 10 week period (tested on age 10-18)		Not known	The same specific peer problem situation was given to both children and parents to rate coping strategies against Children's version used with pupils.

Measure	Details	No. of items	Reliability		Internal consistency	Studies used in
			Test-retest	Inter-rater		
						Parents completed adolescent version to report on their children
Social/Emotional Measures						
Social Skills Questionnaire, Student Form, (SSQSF) Elementary Level (Gresham & Elliot, 1990)	Questionnaire. Assesses cooperation, assertion, self-control and empathy.	30			Lithuania sample: 0.7-0.81 Denmark samples: 0.35-0.62	Mishara & Ystgaard (2006)
Social Skills Questionnaire, Teacher Form, SSQTF, Elementary Level (Gresham & Elliot, 1990)	Questionnaire assessing 3 areas: cooperation, assertion and self-control. Behaviours rated based on their frequency on a Likert-scale. Based on teacher observations.	30		Mishara & Ystgaard (2006), Lithuania sample: 0.4-0.72 ($p < .001$)	Lithuania sample: 0.77-0.8 Denmark sample: 0.81-0.91	Mishara & Ystgaard (2006)
Social Skills Improvement System Rating Scales (SSIS) (Gresham & Elliot, 2008)	4-point Likert scale used to rate the frequency of behaviours: never, seldom, often or almost always. The importance of behaviours are rated as: not important, important, or critical.	Parent version 79 Teacher version 83	Parent version .72-.88 Teacher version 0.68-0.92	Parent version 0.37-0.69 Teacher version 0.36-0.69	Parent version .94-.96 Teacher version 0.94-0.97 Subscale internal consistency 0.74-0.95 (parent version) and 0.75-0.95 (teacher version)	Rodker (2013) used parent & teacher versions (Gresham & Elliot, 2008)

Measure	Details	No. of items	Reliability		Internal consistency	Studies used in
			Test-retest	Inter-rater		
Socio-emotional profile (Dumas et al., 1997)	general academic skills with a score from 1 (lowest 10%) to 5 (highest 10%). Measures social competencies and adaptation problems on a 6-point scale from never to always.	80			Dufour et al. (2011) 0.90	Dufour et al. (2011)
Social support scale for children (Harter, 1985)	Measure perceived social support from peers in class, intimate friends, parents and teachers. Uses 4-point scales.	24			0.66	Dufour et al. (2011)
Emotional literacy checklist (Faupel, 2003)	Questionnaire using a four-point Likert scale. Gives scores for 5 subscales: self-awareness, self-regulation, motivation, empathy and social skills.	20			0.94 Faupel (2003)	Clarke et al. (2014)
NEPSY-II	Developmental neuropsychological evaluation measure. 32 subtests across six domains of functioning. The Social perception domain has 2 subtests: Affect Recognition and Theory of Mind. These were the only subtests used in the reviewed studies.		Affect Recognition .46	Theory of mind subtest 99%.	0.91 Clarke et al. (2014) Affect Recognition 0.67	Rodker (2013) used subtests
			Theory of Mind 0.84		Theory of Mind .84 (Korkman, Kirk & Kemp, 2007)	
Behavioural Measures						
Strengths and Difficulties Questionnaire. Parent and teacher versions. (Goodman, 1997)	Questionnaire measure of emotional and behavioural functioning. Subscale scores: emotional symptoms, hyperactivity/inattention, conduct problems, prosocial behaviour and peer relationship problems. Response scale gives three options: Not True, Somewhat True and Certainly True. English version: ages 4-16. Norwegian version: ages 3–16 years.	25		English version 0.62 (after 4-6 months) Norwegian version not known	English version 0.73 (Goodman, 2001) Holen et al. (2012) Norwegian version not known Clarke et al (2014) 0.76.	Clarke et al. (2014) Holen et al. (2012)

Coping skills measures allowed for multi-source and multi methods of measurement, which is beneficial because it allows for triangulation of the data, which is reflected in WoE A and B. However, these measures varied in their reliability; although the Schoolagers Coping inventory has good internal reliability, the observation form developed based on this measure was used either without reporting its reliability or with the reliability lower than 0.7. Additionally the Kidcope questionnaire was inferior in terms of its reliability, as no internal consistency statistic was reported due to no factor structure being derived. Previous research suggested an exploratory approach to determine the number of factors the Kidcope data fits in each study, as it varies dependent on the age group and stressful situation utilised during the questionnaire (Spirito, 1996).

Social and emotional skills were assessed using a range of measures with no measure being used across studies. However, some reported subscales were similar across measures, allowing for comparison. The Social Skills Questionnaire (Gresham & Elliot, 1990) and Social Skills Improvement System Rating Scales (Gresham & Elliot, 2008) both allowed for multi-source information to be collected through their different versions. As did the Strengths & Difficulties Questionnaire (the only behavioural measure used across studies).

The only direct assessment measure used was the NEPSY-II, which assessed Theory of Mind and Affect Recognition. This may be beneficial as children were young (between ages 5 and 8 years of age) and understanding of questionnaire measures may have been hard to ascertain when administering at whole class levels.

The quality of measures contributed to WoE A ratings, which took into account the validity and reliability of measures. For the measurement rating (as part of WoE A), Clarke et al. (2014) was given a medium rating, whereas all other studies received a low rating, indicating its superiority in regards to this aspect of the study design. Measures were also considered within WoE B; an increased number of sources of measurement, use of pre, post and follow up measures, and use of both active and control comparison groups, were given highest ratings because these aspects of the study design allow for the most accurate assessment of the intervention's effectiveness, as previously discussed. Due to all studies receiving WoE B ratings of between 1.5 and 2.4, they are all considered as 'medium' for their WoE B, indicating there is limited difference between the studies in regards to this. However, it is worth noting that Clarke et al. (2014) received the highest rating out of the reviewed studies for WoE B. Taking this, and its higher rating of 'measurements' as part of WoE A into account, it can be considered that this study was the most methodologically relevant for this review and findings from this study may be of more value.

Findings

Effect sizes were calculated using Morris's (2007) pretest-posttest-control group effect sizes to allow for comparisons to be made across studies. This calculates the standardised mean difference between intervention and control groups, between pre-test and post-test. This was calculated for all studies except Holen et al. (2014), for which effect sizes were extracted from the text.

Effect sizes are reported for subscales of assessment measures, to allow for more meaningful interpretation. Effect size descriptors are given using Cohen’s (1988) guidance, with the additional descriptor of ‘not practically significant’ for effect sizes less than 0.1, as shown in tables 6, 7 and 8.

Table 6

Effect size descriptors

Effect size	Effect size descriptor
0<0.21.9	Not practically significant
0.2-0.49	Small
0.5-0.79	Medium
0.80-1.0	Large

Table 7
Effect sizes summary

Coping Measures	Outcome measure	Study	Sample size at pre-test	Subscale	Effect size	Effect size descriptor	Overall WoE D	
	Schoolagers Coping Inventory	Dufour et al. (2011)	Means and SDs at pre-test and post-test were not given for either control or intervention groups. Statistical analysis not given. Only reported that there were no effects found.			Not practically significant	Medium	
		Mishara & Ystgaard (2006) Denmark sample	Intervention N=322 Control N=110	no. of strategies x helpfulness	0.36	Small	Medium	
	Observation based on Schoolagers coping Inventory Kidcope (children)	Mishara & Ystgaard (2006) Denmark sample	Intervention N=322			Effect size not calculated for this measure due to no control group		Medium
		Holen et al. (2012)	Intervention N=685 Control N=631	Active/Emotional regulation Withdrawal	0.84 -.010	Large NPS	Medium	
	Kidcope (parents)	Holen et al. (2012)	Parents: Intervention N=661 Control N=617	Active/Emotional regulation Withdrawal	-.380 p<.05 -.058	Small Not practically significant	Medium	
				Completing on behalf of children: Intervention N=685 Control N=631	Oppositional	.186 p<.01 .144		Not practically significant Not practically significant
Social/ Emotional measures	Emotional literacy checklist	Clarke et al. (2014)	222 control 267 intervention	self-awareness	0.38	Small	Medium	
				self-regulation	0.18	Not practically significant		
				Motivation	0.3	Small		
				Empathy	0.09	Not practically significant		
				social skills	0.17	Not practically significant		
	Social emotional profile	Dufour et al. (2011)	Intervention N=310 Control N=303	resistant-cooperative	0.19	Not practically significant	Medium	

Outcome measure	Study	Sample size at pre-test	Subscale	Effect size	Effect size descriptor	Overall WoE D
Social Support scale	Dufour et al. (2011)	Intervention N=310 Control N=303	dependent-autonomous	0.16	Not practically significant	Medium
			interiorized problems	0.20	Small	
SSQTF teachers	Mishara & Ystgaard (2006) Denmark sample	Intervention N=322 Control N=110	Support from classmates	0.43	Not practically significant	Medium
			cooperation	-0.032	Not practically significant	
SSQTF students	Mishara & Ystgaard (2006) Denmark sample	Intervention N=322 Control N=110	Assertion	0.05	Not practically significant	Medium
			self-control	0.56	Medium	
SSQTF teachers	Mishara & Ystgaard (2006) Lithuanian sample	Intervention N=314 Control N=104	cooperation	0.04	Not practically significant	Medium
			Assertion	-0.04	Not practically significant	
SSQTF students	Mishara & Ystgaard (2006) Lithuanian sample	Intervention N=314 Control N=104	self-control	-0.15	Not practically significant	Medium
			Empathy	0.43	Small	
SSQTF teachers	Mishara & Ystgaard (2006) Lithuanian sample	Intervention N=314 Control N=104	Cooperation	0.17	Not practically significant	Medium
			Assertion	0.58	Medium	
SSQTF students	Mishara & Ystgaard (2006) Lithuanian sample	Intervention N=314 Control N=104	self-control	0.60	Medium	Medium
			Cooperation	0.02	Not practically significant	
			Assertion	0.26	Small	
			self-control	0.28	Small	
			Empathy	0.21	Small	

Outcome measure	Study	Sample size at pre-test	Subscale	Effect size	Effect size descriptor	Overall WoE D	
Social Skills Improvement System Rating Scales (parent version)	Rodker (2013)	Intervention N=65 Control N=60				Not practically significant	Low
			Social Skills	-0.11	Not practically significant		
			Communication	-0.04	Not practically significant		
			Cooperation	-0.12	Not practically significant		
			Assertion	-0.32	Small		
			Responsibility	0.04	Not practically significant		
			Empathy	-0.24	Small		
			Engagement	0.15	Not practically significant		
			Self-Control	-0.19	Not practically significant		
			Problem Behaviours	-0.23	Small		
			Externalizing	-0.13	Not practically significant		
			Bullying	-0.10	Not practically significant		
			Hyperactivity/Inattention	-0.06	Not practically significant		
			Internalizing	-0.38	Small		
Social Skills Improvement System Rating Scales (Teacher version)	Rodker (2013)	Intervention N=65 Control N=60	Autism (Social Skills)	-0.02	Not practically significant	Low	
			Autism (Behaviour Problems)	-0.21	Small		
			Social Skills	-0.1	Not practically significant		
			Communication	0.15	Not practically significant		

Outcome measure	Study	Sample size at pre-test	Subscale	Effect size	Effect size descriptor	Overall WoE D	
			Cooperation	-0.07	Not practically significant		
			Assertion	-0.17	Not practically significant		
			Responsibility	-0.18	Not practically significant		
			Empathy	-0.13	Not practically significant		
			Engagement	-0.07	Not practically significant		
			Self-Control	0.01	Not practically significant		
			Problem Behaviours	0.25	Small		
			Externalizing	0.38	Small		
			Bullying	0.31	Small		
			Hyperactivity/Inattention	0.18	Not practically significant		
			Internalizing	0.05	Not practically significant		
			Autism (Social Skills)	-0.11	Not practically significant		
			Autism (Behaviour Problems)	-0.02	Not practically significant		
			Academic Competence	-0.06	Not practically significant		
NEPSY-II	Rodker (2013)	Intervention N=65 Control N=60	Affect Recognition	0.14	Not practically significant	Low	
			Theory of Mind	-0.45	Small		
Behavioural Measures	SDQ	Clarke et al. (2014)	222 control 267intervention	emotional symptoms	-0.16	Not practically significant	Medium
				conduct problems	0.11	Not practically significant	
				Hyperactivity peer relationship problems	-0.11	Not practically significant	
					0.03	Not practically significant	

Outcome measure	Study	Sample size at pre-test	Subscale	Effect size	Effect size descriptor	Overall WoE D
SDQ (parents)	Holen et al. (2012)	Parents: Intervention N=661 Control N=617	prosocial behaviour	0.1	Not practically significant	Medium
			Pro—social behaviour	-.031	Not practically significant	
			Hyperactivity/ inattention	-.032	Not practically significant	
			Completing on behalf of children:	.025	Not practically significant	
			Intervention N=685	-.024	Not practically significant	
			Control N=631	-.040	Not practically significant	
SDQ (teachers)	Holen et al. (2012)	Teachers: Intervention N:685 Control N=625	Pro—social behaviour	.176	Not practically significant	Medium
			Hyperactivity/ inattention	-.098	Not practically significant	
			Completing on behalf of children:	.117	Not practically significant	
			Intervention N=685	-.057	Not practically significant	
			Control N=631	-.047	Not practically significant	
				-.146 p<.05	Not practically significant	

Table 8
Effect sizes at follow up

	Outcome measure	Study	Sample size at pre-test	Subscale	Effect size	Effect size descriptor	Overall WoE D
Social/ Emotional measures	Emotional literacy checklist	Clarke et al. (2014)	222 control 267 intervention	self-awareness	0.01	No effect	Medium
				self-regulation	0.00	No effect	
				Motivation	0.01	No effect	
				Empathy	0.00	No effect	
				social skills	-0.01	No effect	
Behavioural Measures	SDQ	Clarke et al. (2014)	222 control 267intervention	emotional symptoms	0.00	No effect	Medium
				conduct problems	0.01	No effect	
				Hyperactivity	0.00	No effect	
				peer relationship problems	0.00	No effect	
				prosocial behaviour	0.00	No effect	

Only one large effect size was found, which was for use of emotional regulation as a coping strategy as measured by the Kidcope questionnaire. However, concerns around this measure's reliability calls into question whether this would be a consistent finding if the measure were used across studies. Mishara and Ystgaard's (2006) study found an increase in helpful coping strategies (small effect) in their Denmark sample and Holen et al. (2011) found decreased oppositional coping strategies (small effect) as rated by children. Other coping skills measures were not practically significant, although these were predominantly also from the Kidcope questionnaire, so again this may reflect the unreliability of the measure.

Mishara and Ystgaard (2006) found medium effect sizes for an increase in self-control in their Denmark sample and both increases in assertion and self-control in their Lithuanian sample, as measured by the Social Skills Questionnaire Teachers Form. The student version of this questionnaire indicated small effect sizes for increases in these subscales (assertion and self-control) in the Lithuanian sample, although these were not found in the Denmark sample. Additionally, student forms in both samples found small effect sizes for increases in empathy. The increased effect size in teacher ratings in comparison to student ratings for these subscales may reflect children being less self-aware of these abilities. Having both teacher and student agreement allows for a stronger degree of certainty over effects on assertion and self-control. Further to this, improvements in empathy and assertion were found by Rodker (2013), increasing certainty of these being real effects, as effects were found across multiple studies and measures.

Additional effects found across social and emotional measures were all small in size: increases in self-awareness, motivation, theory of mind, support from teachers, and decreases in problem behaviours, internalizing, externalising and bullying. These effects were all in the expected direction, with positive aspects of social and emotional functioning increasing, and negative aspects decreasing. These were all found in one study, using one measure.

Just over seventy percent of subscale measure were found to be not practically significant, indicating there is a lack of consistency of findings. However, a number of these were from Rodker's (2013) study, which received a low WoE A rating, calling into question the methodological rigour of the study and its appropriateness to the current review, as it had no follow up, nothing to establish equivalence of groups and nothing done to ensure fidelity. Therefore, the quantity of not practically significant results should not necessarily lead to the intervention being discarded as a potential intervention to support social and emotional functioning.

No effects were found at follow up, calling into question the longevity of benefits. As the follow-up was conducted by one of the studies rated as medium for its weight of evidence (which was the highest rating in this review), it raises additional concern that long-term effects are not gained from this intervention, as it is less likely that the lack of follow-up effects is due to the study being poorly designed.

Conclusions and recommendations

Overall, the Zippy's Friends intervention shows some promise as a universal intervention to improve children's social and emotional wellbeing, as small-medium effect sizes have been found across a range of areas pertaining to social and emotional functioning. Although there was a considerable amount of results that were not practically significant, these should not result in the dismissal of Zippy's friends as an effective intervention because no studies received high ratings for methodological quality or relevance. Therefore, the lack of significant findings may be due to poor research design rather than a lack of true effects, which calls for further research.

Due to the wide-ranging effects, practitioners should be cautious about recommending Zippy's Friends for a specific 'need', such as the 'need' to develop appropriate coping skills when children are displaying internalising or externalising coping strategies. Previous adaptations to the programme were formulated to improve the programme's impact on coping skills, which was the intended use of the programme. Small to large effect sizes were found on coping measures (although not all reliable measures), which indicates amendments have led to benefits in this area. However, the majority of measures used and effect sizes found were related to social and emotional functioning. It would therefore be beneficial to acknowledge these wide-ranging potential outcomes and recommend this intervention as addressing this range of skills, rather than indicating such a focus on coping. This is not necessarily a downfall in a universal intervention; it is not only targeting children lacking coping skills, so these other social/emotional benefits should be advantageous for all. However, it is important to

acknowledge these ranging effects, as practitioners may not realise the potential benefits being gained if they are only expecting these benefits to pertain to coping skills.

Content from this universal intervention could potentially be drawn upon in personal, social and health education lessons to incorporate knowledge and strategies into a whole-school approach, to complement its implementation. In addition, further consideration of how parents may be involved may strengthen effects found, through additional discussion and modelling of strategies in the home setting.

It should be noted that no study included in this review received a 'high' overall rating for WoE D, reflecting the lack of sufficiently rigorous methodology and approaches suitable to answering this review question. It is therefore necessary for future studies into the Zippy's Friends Programme to implement randomised designs and use reliable measures, in order to more accurately assess the effectiveness of the intervention. Additionally, active control groups, rather than no intervention controls would allow for benefits to be compared against those gained from alternative interventions. This is of particular relevance for practitioners making recommendations; there may be costs associated with different programmes and educators will be required to make judgements of which interventions are most beneficial and if outcomes justify costs.

In terms of further research, it would be beneficial for more studies to be carried out in the U.K in order to ascertain whether the translated version is effective. One reviewed study was carried out in Ireland (Clarke et al., 2014), but this was implemented in disadvantaged schools, so was not representative. However, it would be beneficial to measure teacher perceptions of the intervention and fidelity to its delivery in these

cases, as commitment to implementation of the intervention could be affected by its unfamiliarity in the U.K.; teachers may not expect the intervention to be as useful if they have not heard of it before. It would also be interesting to carry out a longitudinal study to monitor whether Zippy's Friends has any long-term effects on coping with later life events, as this is the theoretical premise the intervention was devised on.

References

- Clarke A. M., & Barry, M. M. Implementing mental health promoting schools. In: Simovska V., McNamara, P. (eds). *Schools for Health and Sustainability*. Dordrecht, the Netherlands: Springer, in press.
- Clarke, A. M., Bunting, B., & Barry, M. M. (2014). Evaluating the Implementation of a School-Based Emotional Well-Being Programme: A Cluster Randomized Controlled Trial of Zippy's Friends for Children in Disadvantaged Primary Schools. *Health Education Research*, 29(5), 786-798.
- Clarke, A. M. (2011). An evaluation of the Zippy's Friends emotional wellbeing programme for primary schools in Ireland. PhD Thesis. Ireland: Discipline of Health Promotion, National University of Ireland Galway.
- Clarke, A. M., O'Sullivan M, & Barry, M. M. (2010). Context matters in programme implementation. *HealthEduc*. 110, 273–93.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. USA: Lawrence Erlbaum Associates.
- Cohen, J. (1992). Quantitative Methods in Psychology. A Power Primer. *Psychological Bulletin*, 112(1), 155–159.
- Cottrell, D. (2002). Body of evidence: Towards evidence based practice. *Young Minds*, 58, 34–37.
- Department for Education (DfE) and Department of Health (DH). (2014). *SEND code of practice: 0 to 25 years*. London: The Stationery Office.
- Department for Education and Skills (DfES). (2001). *Promoting children's mental health within early years and school settings*. London: DfES.

Dufour, S., Denoncourt, J., & Mishara, B. L. (2011). Improving Children's Adaptation: New Evidence Regarding the Effectiveness of Zippy's Friends, a School Mental Health Promotion Program, *Advances in School Mental Health Promotion*, 4(3), 18-28.

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(1), 405–432.

Ermisch, J. (2008). Origins of social immobility and inequality: parenting and early child development. *National Institute Economic Review*, 205(1), 62-7.

Farrell, P., Woods, K., Lewis, S., Rooney, S., Squires, G., & Connor, M. O. (2006). A *Review of the Functions and Contribution of Educational Psychologists in England and Wales in light of "Every Child Matters: Change for Children."* (Research report. Great Britain. Department for Education and Skills; no. 792). Nottingham: DfES.

Faupel, A. (2003). *Emotional Literacy: Assessment and Intervention—Ages 7 to 11*. London: nferNelson.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research papers in education*, 22(2), 213-228.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 38(5), 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.
- Greenberg, M. T., Weissberg, R. P. O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing School-Based Prevention and Youth Development Through Coordinated Social, Emotional, and Academic Learning. *American psychologist, 58*(6/7), 466–474.
- Gresham, F. M., & Elliot, S. N. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.
- Gresham, M. F., & Elliott, S. N. (2008). *Social Skills Improvement System rating scales manual*, Minneapolis, MN: Pearson.
- Hampel, P. & Petermann, F. (2006) Perceived stress, coping, and adjustment in adolescents. *Journal of Adolescent Health, 38*(4) 409–15.
- Holen, S., Waaktaar, T., Lervåg, A., & Ystgaard, M. (2012). The effectiveness of a universal school-based programme on coping and mental health: a randomised, controlled study of Zippy's Friends, *Educational Psychology, 32*(5), 657-677.
- Korkman, M., Kirk, U., & Kemp, S. L. (2007). NEPSY-II. Clinical and interpretative manual. San Antonio, TX: Psychological Corporation.
- Kratochwill, T. R. (2003). Evidence-Based Practice: Promoting Evidence-Based Interventions in School Psychology. *School Psychology Quarterly, 18*(4), 389-408.
- Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal and coping*. New York, NY: Springer.

- Mishara, B. L., & Ystgaard, M. (2000). Exploring the potential of primary prevention: evaluation of the Befrienders International Reaching Young Europe Pilot Programme in Denmark. *Crisis, 21*(1), 4–7.
- Mishara, B., & 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*(1), 110-123.
- Morris, S. B. (2007). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods, 11*(2), 364–386.
- Nes, L. S., & Segerstrom, S. C. (2006). Dispositional optimism and coping: A meta-analytic review. *Personality and social psychology review, 10*(3), 235-251.
- Power, C. & Matthews, S. (1997). Origins of health inequalities in a national population sample. *Lancet, 350*(9091), 1584-9.
- Rodker, J. D. (2013). Promoting Social-Emotional Development of Children During Kindergarten: A Zippy's Friends Program Evaluation. (unpublished doctoral thesis). Pace University, New York, America.
- Ryan-Wenger, N. (1990). Development and psychometric properties of the Schoolagers Coping Strategies Inventory. *Nursing Research, 39*(6), 344–349.
- Scheier, M. F., & Carver, C. S. (1987). Dispositional optimism and physical well-being: The influence of generalized outcome expectancies on health. *Journal of Personality, 55*, 169-210.
- Segal, J. (1983). Translating stress and coping research into public information and education. In M. Kessler & S. E. Goldston (Eds.), *A decade of progress in primary prevention*. Lebanon, New Hampshire: University Press of New England.

Spirito, A., Stark, L. J., & Williams, C. (1988). Development of a brief coping checklist for use with pediatric populations. *Journal of Pediatric Psychology, 13*(4), 555–574.

Spirito, A. (1996). Commentary: Pitfalls in the use of brief screening measures of coping. *Journal of Pediatric Psychology, 21*(4), 573–575.

Appendices

Table 1

Appendices content list

Appendix	Containing	Page
A	Modules, sessions and goals of Zippy’s Friends intervention	44
B	Excluded studies with rationale	46
C	Table 4: Mapping the field. Key information about the studies included for full review.	49
D	Amendments made to Kratochwill (2003) coding protocol with rationale	53
E	Coding protocols	55
F	Weight of Evidence A criteria and summary table	69
G	Weight of Evidence B criteria and summary table	74
H	Weight of Evidence C criteria and summary table	76
I	Weight of Evidence D averaging criteria	77

Appendix A

Table 1

**Modules, sessions and goals of Zippy’s Friends intervention
Taken from Mishara and Ygstaard (2006)**

Module	Sessions
Module 1: Feelings Goal: To improve children’s abilities to recognize negative feelings and to identify coping strategies to deal with these feelings	SESSION 1: FEELING SAD–FEELING HAPPY To improve children’s abilities to recognize feeling sad and to identify ways to cope with feeling sad SESSION 2: FEELING ANGRY OR ANNOYED To improve children’s abilities to recognize feelings of anger and annoyance and to identify ways to cope with these feelings SESSION 3: FEELING JEALOUS To improve children’s abilities to identify feelings of jealousy and to learn ways to cope with this feeling SESSION 4: FEELING NERVOUS To improve children’s abilities to recognize feeling nervous and to identify ways to cope with this feeling
Module 2: Communication Goal: To improve children’s abilities to communicate their feelings	SESSION 1: IMPROVING COMMUNICATIONS To improve children’s abilities to recognize characteristics of effective and ineffective communications SESSION 2: LISTENING To improve children’s abilities to listen SESSION 3: WHO CAN HELP US? To improve children’s abilities to ask for help SESSION 4: SAYING WHAT YOUWANT TO SAY To improve children’s abilities to say what they want to say
Module 3: Making and breaking relationships Goal: To improve children’s abilities to make friends and to cope with rejection and loneliness	SESSION 1: HOW TO KEEP A FRIEND To improve children’s abilities to recognize how to keep their friends SESSION 2: DEALING WITH LONELINESS AND REJECTION To improve children’s abilities to cope with loneliness and rejection SESSION 3: HOW TO RESOLVE CONFLICTS WITH FRIENDS To improve children’s abilities to resolve conflicts with friends SESSION 4: HOW TO MAKE FRIENDS To improve children’s abilities to make friends
Module 4: Conflict resolution Goal: To improve children’s abilities to resolve conflicts	SESSION 1: HOW TO RECOGNISE GOOD SOLUTIONS To increase children’s abilities to recognize characteristics of a good solution SESSION 2: BULLYING To improve children’s abilities to deal with situations involving bullying SESSION 3: SOLVING PROBLEMS To improve children’s abilities to resolve conflicts SESSION 4: HELPING OTHERS RESOLVE CONFLICTS To improve children’s abilities to help others resolve conflicts

Module 5: Dealing with change and loss Goal: To improve children's abilities to cope with change and loss	SESSION 1: CHANGE AND LOSS ARE PART OF LIFE To increase the children's understanding that change and loss are part of normal everyday experiences SESSION 2: COPING WITH DEATH To increase the children's understanding that death is a normal part of life; and to improve children's abilities to cope with situations involving grief and loss SESSION 3: VISIT TO A GRAVEYARD To improve children's abilities talk about death and loss SESSION 4: BENEFITS OF CHANGE AND LOSS To improve children's understanding that change and loss often have positive effects
Module 6: We cope Goal: To improve children's abilities to use a variety of coping strategies in different situations	SESSION 1: DIFFERENT WAYS TO COPE To improve children's abilities to use different coping strategies SESSION 2: HOW TO HELP OTHERS To improve children's abilities to help others cope with different situations SESSION 3: ADAPTING TO NEW SITUATIONS To improve children's abilities to generalize their coping skills to new situations SESSION 4: CELEBRATING TOGETHER To review what we have learned during this program and to celebrate Together

Appendix B

Table 2

Excluded studies with rationale

Excluded Study	Rationale
<p>Monkevicienė, O., Mishara, B., & Dufour, L. (2006). Effects of the Zippy's Friends Programme on Children's Coping Abilities During the Transition from Kindergarten to Elementary School. <i>Early Childhood Education Journal</i>, 34(1), 53-60.</p>	<p>Criteria 6: Study design Pre-intervention data was not provided, as Zippy's Friends was not implemented as part of the study; the intervention group were children who had already received the intervention during pre-school</p>
<p>Clarke, A. M. (2001). An evaluation of Zippy's Friends, an emotional wellbeing programme for children in primary schools. (Unpublished Doctoral Thesis). National University of Ireland Galway, Ireland.</p>	<p>Criteria 8: Data Data used in another study used in the review (Clarke, Bunting & Barry, 2004). This study was discounted, due to being a thesis, as the other study using the same data was published in a peer reviewed journal</p>

Excluded Study	Rationale
	<p>and would therefore be less likely to contain errors, so was included in preference</p>
<p>Clarke, A. M., O'Sullivan, M., & Barry, M. M. (2010). Context matters in programme implementation, <i>Health Education, 110</i>(4), 273 – 293.</p>	<p>Criteria 6: study design Case studies</p>
<p>Clarke, A., Sixsmith, J., & Barry, M. (2015). Evaluating the implementation of an emotional wellbeing programme for primary school children using participatory approaches. <i>Health Education Journal, 74</i>(5), 578-593.</p>	<p>Criteria 4: outcomes Participatory approaches used; qualitative data</p>
<p>Rimkevičienė, V., Mauragienė, V. (2016). Tevu požiuris i vaiku išgyvenamus socialinius sunkumus ir jų poveikio budus adaptacijos pradinėje mokykloje laikotarpiu, baigus programą „Žipio draugai“. <i>Pedagogika, 123</i>(3), 172-186.</p>	<p>Criteria 7: Language. Study is written in Lithuanian</p>
<p>Clarke, A. M., & Barry, M. M. Implementing mental health promoting schools. Implementing mental health promoting schools. In: Simovska V., McNamara, P. (eds). <i>Schools for Health and Sustainability</i>. Dordrecht, the Netherlands: Springer, in press.</p>	<p>Criteria 1: Publication type Book chapter</p>
<p>Monkevičienė, O. (2014). Ankstyvosios prevencijos programos „Žipio draugai“ poveikis vaikų socialinės</p>	<p>Criteria 7: Language. Study is written in</p>

Excluded Study	Rationale
<p>kompetencijos ugdymuisi bei saugios emocinés aplinkos ugdymo įstaigose kūrimui. <i>Pedagogika</i>, 116(4), 71-93.</p>	Lithuanian
<p>Holen, S., Waaktaar, T., Lervåg, A., Ystgaard, M. (2013). Implementing a Universal Stress Management Program for Young School Children: Are there Classroom, or Academic Effects? <i>Scandinavian Journal of Educational Research</i>. 57(4), pp. 420-444</p>	<p>Criteria 4: Outcomes Not looking at social/emotional/mental health outcomes.</p>
<p>Clarke, A., Barry, M. (2012). An evaluation of the Zippy's Friends programme in disadvantaged primary schools in Ireland. <i>Psychology & Health</i>, 27(1), 21-22.</p>	<p>Criteria 1: Publication type. Meeting abstract</p>
<p>Bale, C., & Mishara, B. (2004) Developing an International Mental Health Promotion Programme for Young Children. <i>International Journal of Mental Health Promotion</i>, 6(2), 2004, pp. 12-16.</p>	<p>Criteria 6: Study design Article detailed programme, but it was not an experimental study implementing Zippy's Friends</p>

Appendix C

Table 1
Mapping the field. Key information about the studies included for full review.

Study	Location	Design	Delivery/ training/ support	Participant details	Groups	Outcomes measure	Key findings
Clarke et al. (2012)	Ireland	Clustered, randomised controlled trial. Randomisation occurred at school level. Pre & post-tests and 12 month follow up	2 day training from Health Promotion Specialists. Ongoing support from Health Promotion Specialists (school visits and an interim group meeting)	Children aged 7-8 years Schools given Disadvantaged status from the Department of Education and Skills Intervention Type I N=276 132 boys, 135 girls Intervention Type II N=277 134 boys, 117 girls Control group N=222 children 113 boys, 99 girls	Control: PSHE in typical curriculum Type I intervention: Zippy's Friends implemented faithfully Intervention Type II: Zippy's Friends as a resource along with PSHE elements as desired	Emotional Literacy checklist Strengths and difficulties questionnaire	Significant direct effect of the intervention on the intervention group's: Self-awareness Estimate=0.39, SE=0.057, C.R.=6.875, P<0.001), Std Est=0.351 Self-regulation Estimate=0.220, SE=0.083, C.R.=2.66, P<0.01) Std Est=0.122 Motivation Estimate=0.215, SE=0.058, C.R.=3.691, P<0.001), Std Est=0.133 Social Skills score Estimate=0.215, SE=0.058, C.R.=3.691, P<0.001), Std Est=0.124. Sustained positive effect on emotional literacy at 12 month follow up Intervention had no significant effect on the intervention group's emotional symptoms Estimate=0.068, SE=0.51, C.R.=1.32, P=0.188), Std, Est=0.109]; hyperactivity Estimate=0.046, SE=0.077, C.R.=0.604, P=0.546), Std, Est=0.037 peer relationship problems Estimate=0.054, SE=0.047, C.R.=1.141, P=0.254), Std, Est¼ 0.094 prosocial behaviour Estimate=0.058, SE=0.054, C.R.=1.075, P=0.282), Std, Est=0.076 Conduct problems significantly decreased in the control group Estimate=-0.132, SE=0.059, C.R.=-2.242.003, P<0.05, Std, Est=0.141. This was not sustained at 12 month follow up. No effects found at 12 month follow up for

those that had not had an effect at post intervention.

Study	Location	Design	Delivery/training/support	Participant details	Groups	Outcomes measure	Key findings
Dufour et al. (2011)	Canada, Quebec	Quasi-experimental design. Pre and post tests	1 day training Half a day follow up meeting during implementation	Average age 6.72 years Intervention: N=310 141 girls, 169 boys Control: N=303 148 girls, 155 boys 35 teachers 29 had >8yrs teaching experience. 28 had worked for at least 3 years in their current school. SES of teachers and children unknown	Control: no intervention	Schoolagers coping strategies Inventory (Ryan-Wenger, 1990)	Between groups comparison showed the intervention group were more cooperative towards adults $F=13.30, p \leq 0.0001, df = 1,559, \eta^2p = 0.023$ Had increased autonomy $F = 9.37, p = 0.002, df = 1,558; \eta^2p = 0.017$ Had fewer internalised problems $F = 12.05, p = 0.001, df = 1,560; \eta^2p = 0.021$ The intervention group had an increase in perceived teacher support $F=7.43, p = 0.007, df = 1,240; \eta^2p=0.03$ Control group perceived increase from classmates, but intervention group did not $F = 5.06, p < 0.025, df=1,240, \eta^2p=0.02$ No effects found in relation to coping strategies or their perceived effectiveness
Holen et al. (2012)	Norway	Schools matched in pairs (on SES profile, percentage of special teaching and ethnic minority backgrounds) and randomly	2 days of training. 3 counselling sessions lasting one day each	1297 participated at pre-test Children aged 7-8 years 49.3% girls 75.9% lived with both mother and father At least one parent	Control: no intervention	Kidcope questionnaire (Spirito, Stark, and Williams, 1988) Extended Norwegian version of the Strengths and Difficulties Questionnaire. Parent and teacher versions.	children reported a significant reduction (Cohen's $d = .380$) in oppositional coping strategies Parents reported a significant increase in active strategies (Cohen's $d = .186$). SDQ subscales were not found to have significant changes across pre/post tests Teachers reported a significant decrease in impact score (Cohen's $d = .146$). Girls used significantly less oppositional coping, based on self-report (Cohen's $d = .551$)

assigned to intervention or control.

had completed high school for 85.7% and 61.6% received higher education

(Goodman, 1997)

significant increase in active coping strategies, as reported by parents (Cohen's $d = .258$)

Study	Location	Design	Delivery/training/support	Participant details	Groups	Outcomes measure	Key findings
Mishara & Ystgaard (2006)	Denmark & Lithuania	Quasi-experimental design. Pre & post-tests	2 day training Coordinator in each country 'met regularly' with teachers for ongoing support	Denmark Intervention: N=322 160 boys and 162 girls Control group: N=110 53 boys and 57 girls Lithuania Experimental group: N=314 171 boys and 143 girls Control group N=104 52 boys and 52 girls	Control: no intervention In Denmark they had no control group for behaviour measure	SSQTF, Social Skills Questionnaire, Teacher Form, Elementary Level SSQTF, Social Skills Questionnaire, Student Form, Elementary Level Teacher observation based on the Schoolagers coping strategies Inventory (Ryan-Wenger, 1990) Schoolagers coping strategies Inventory (Ryan-Wenger, 1990)	Between group comparison showed a significant pre-post difference for social skills in Lithuania $F(3,400) = 16.63, p < .001, \eta^2p = .110$ and Denmark $F(3,368) = 10.69, p < .001, \eta^2p = .080$ In Lithuania, this improvement was in assertion, self-control and cooperation. In Denmark, it related only to self-control. Repeated measures MANOVA indicated significant decreases in problem behaviours. $F=10.40, d.f.=3, 691, p < .001, \eta^2p = .043$. This was in relation to externalising and hyperactivity and combined Lithuania and Denmark samples Repeated measures ANOVA indicated significant increase in coping strategies used for intervention group (based on teacher observations) in Lithuania $F=10.61, d.f.=1,404, p < .001, \eta^2p = .068$ And in Denmark $F(1,394) = 4.51, p < .015, \eta^2p = .008$

Study	Location	Design	Delivery/ training/ support	Participant details	Groups	Outcomes measure	Key findings
Rodker (2013)	America, New Jersey	Pre & post- test (post test data collection occurred 2 months after the programme ended)	Unknown level of training; School were already implementing Zippy's Friends	N=125 (after drop- out of 2) 52 females and 73 males. Intervention group: N=65 28 girls, 37 boys Control group N=60 24 girls, 36 boys	Control group: no intervention	Social Skills Improvement System Rating Scales (SSIS) NEPSY II – Affect recognition & Theory of Mind subscales	Parents rated increased Responsibility which was significant in between group analysis, F (1, 49) = 5.19, $p < .05$, $\eta^2 = .10$ Between group comparisons (ANOVA) did not show significant differences between groups for social skills, problem behaviours, affect recognition or Theory of Mind.

Appendix D

Coding protocols from APA Task Force Coding Protocol by Kratochwill (2003) were used in this review. The amendments to this protocol, along with the rationale are detailed in Table 1.

Table 1

Amendments made to the original Kratochwill (2003) protocol and rationale.

Section Heading	Section removed/modified	Rationale
I. General Characteristics	B7. Coding	Qualitative research
	B.8 Interactive process followed	methods not used
II. Key Features for Coding Studies and Rating Level of Evidence/ Support	C. Primary/Secondary Outcomes Are Statistically Significant	Primary outcomes discussed in current review. Secondary outcomes not relevant to the review question
II. Key Features for Coding Studies and Rating Level of Evidence/ Support	E. Identifiable components	None of the studies included referred to particular aspects/ sessions being specifically necessary to produce change
II. Key Features for Coding Studies and Rating Level of Evidence/ Support	E.1 Evidence for primary outcomes	Required coding from prior outcomes section which was removed

Section Heading	Section	Rationale
	removed/modified	
II. Key Features for Coding Studies and Rating Level of Evidence/ Support	D. Educational/Clinical significance	Studies included in the review are not from clinical populations
II. Key Features for Coding Studies and Rating Level of Evidence/ Support	G. Replication	Study replications would have been analysed and treated as a separate study in relation to the review question
II. Key Features for Coding Studies and Rating Level of Evidence/ Support	H.2 Non-school site	All studies involved school-based intervention
III. Other Descriptive or Supplemental Criteria to Consider	G. Intervention style or orientation	All studies were based on models of coping

Appendix E

Example adapted protocol completed.

Study 1

School- and community-based intervention programs for social and behavioural problems	
Academic intervention programs	
Family and parent intervention programs	
School-wide and classroom-based programs	
Comprehensive and coordinated school health services	

Name of Coder(s):

Date: 03/02/2017

Full Study Reference in APA format:

Clarke, A. M., Bunting, B., & Barry, M. M. (2014). Evaluating the Implementation of a School-Based Emotional Well-Being Programme: A Cluster Randomized Controlled Trial of Zippy's Friends for Children in Disadvantaged Primary Schools. *Health Education Research, 29*(5), 786-798.

Intervention Name (description from study): Zippy's Friends

Study ID Number (Unique Identifier): 1

Type of Publication: (Check one)

Book/monograph	
Journal article	
Book chapter	
Other (specify)	

I. General Characteristics

A. General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)		
A1.1		Completely randomized design
A1.2		Randomized block design (between-subjects variation)
A1.3		Randomized block design (within-subjects variation)
A1.4		Randomized hierarchical design

Randomisation occurred at school level. Randomisation did not occur at an individual level, as children were nested within schools. Schools only included if mixed gender and had disadvantaged status and classes were matched on grade.

A2. Nonrandomized designs (if non-random assignment design, select one of the following)	
A2.1	Nonrandomized design
A2.2	Nonrandomized block design (between-participants variation)
A2.3	Nonrandomized block design (within-participants variation)
A2.4	Nonrandomized hierarchical design
A2.5	Optional coding of Quasi-experimental designs (see Appendix C)

A3. Overall confidence of judgment on how participants were assigned (select one of the following)	
A3.1	Very low (little basis)
A3.2	Low (guess)
A3.3	Moderate (weak inference)
A3.4	High (strong inference)
A3.5	Very high (explicitly stated)
A3.6	N/A
A3.7	Unknown/unable to code

B. Statistical Treatment/Data Analysis (answer B1 through B6)

B1.	Appropriate unit of analysis		Yes		No		N/A
B2.	Familywise error rate controlled		Yes		No		N/A
B3.	Sufficiently large <i>N</i>		Yes		No		N/A

Statistical Test: Structural equation modelling. Chi square test (χ^2) used to evaluate the fit of the model. Comparative Fit index (CFI) and root-mean-square error of approximation (RMSEA) also used.

Level: Individual participant level and in terms of the occasions

ES: small

N required: 393 per group

This study had enough participants to detect a medium effect size (more than 64 per group)

B4. Total size of sample (start of the study): 766 students and 52 teachers

B5. Intervention group sample size: type 1 intervention (asked to implement Zippy's Friends faithfully) 267 children and 18 teachers, type 2 intervention (asked to use Zippy's Friends as a resource) 277 children and 18 teachers

B6. Control group sample size: 222 children and 16 teachers

Rating for Statistical Analysis (select 0, 1, 2, or 3)

0	
1	
2	
3	

Appropriate analysis and unit of measurement and used enough participants to detect a medium effect size (as stated in amended WoE A criteria)

C. Type of Program (select one)

C1		Universal prevention program
C2		Selective prevention program
C3		Targeted prevention program
C4		Intervention/ Treatment
C5		Unknown

D. Stage of the Program (select one)

D1		Model/demonstration programs
D2		Early stage programs
D3		Established/institutionalized programs
D4		Unknown

Established, however, there have been more recent changes due to prior research finding effects on social skills and not anticipated coping skills.

E. Concurrent or Historical Intervention Exposure (select one)

E1		Current exposure
E2		Prior exposure
E3		Unknown

II. Key Features for Coding Studies and Rating Level of Evidence/ Support

3=Strong Evidence

2=Promising Evidence

1=Weak Evidence

0=No Evidence

A. Measurement (answer A1 through A4)

A1. Use of outcome measures that produce reliable scores for the majority of primary outcomes. The table for Primary/Secondary Outcomes Statistically Significant allows for listing separate outcomes and will facilitate decision making regarding measurement (select one of the following)

A1.1	<input checked="" type="checkbox"/>	Yes
A1.2	<input type="checkbox"/>	No
A1.3	<input type="checkbox"/>	Unknown/unable to code

SDQ – Cronbach α : .73, cross-informant correlation was found to have a mean 0.34, and retest stability after 4 to 6 months had a mean of 0.62 Goodman (2001)

Emotional Literacy Checklist – 0.94 Cronbach α in prior research (Faupel, 2003) and 0.91 in this study.

A2. Multi-method (select one of the following)

A2.1	<input type="checkbox"/>	Yes
A2.2	<input checked="" type="checkbox"/>	No
A2.3	<input type="checkbox"/>	N/A
A2.4	<input type="checkbox"/>	Unknown/unable to code

Questionnaire measures used – Emotional Literacy Checklist and Strengths and Difficulties Questionnaire

A3. Multi-source (select one of the following)

A3.1	<input type="checkbox"/>	Yes
A3.2	<input checked="" type="checkbox"/>	No
A3.3	<input type="checkbox"/>	N/A
A3.4	<input type="checkbox"/>	Unknown/unable to code

Questionnaires both completed by teachers.

A4. Validity of measures reported (select one of the following)

A5.1	<input type="checkbox"/>	Yes validated with specific target group
A5.2	<input type="checkbox"/>	In part, validated for general population only
A5.3	<input type="checkbox"/>	No
A5.4	<input checked="" type="checkbox"/>	Unknown/unable to code

SDQ validated by comparing with psychiatric diagnoses (Goodman, 2001). Emotional Literacy Checklist validity unknown.

Rating for Measurement (select 0, 1, 2, or 3)

0	<input type="checkbox"/>
1	<input type="checkbox"/>
2	<input checked="" type="checkbox"/>
3	<input type="checkbox"/>

B. Comparison Group

B1. Type of Comparison Group (select one of the following)

B1.1		Typical contact
B1.2		Typical contact (other) specify:
B1.3		Attention placebo
B1.4		Intervention elements placebo
B1.5		Alternative intervention
B1.6		Pharmacotherapy
B1.7		No intervention
B1.8		Wait list/delayed intervention
B1.9		Minimal contact
B1.10		Unable to identify comparison group

One no intervention control group and another comparison group where intervention was given and told to be used as a resource.

Rating for Comparison Group (select 0, 1, 2, or 3):

0	
1	
2	
3	

Both active comparison and no intervention comparison groups used.

Equivalent mortality and low attrition.

- Type 1 Intervention: 267 children allocated. 0 lost at baseline, 48 children lost at post intervention (18% attrition) and 14 at follow up (5%)
- Type 2 Intervention 277 children allocated. 26 lost at baseline (9% attrition), 23 children lost at post intervention (8%) and 6 at follow up (2% attrition overall)
- Type 2 Intervention 222 children allocated. 10 lost at baseline (5% attrition), 23 children lost at post intervention (10% attrition) and 4 at follow up (2%)
Percentages calculated using total initially allocated to each group and rounding to the nearest whole number.

No significant difference across groups for gender, location (urban/rural) and whether multi-grade classes.

No counterbalancing of change agents reported – teachers delivering the intervention would have varied across classes/schools.

B2. Overall confidence rating in judgment of type of comparison group (select one of the following)

B2.1		Very low (little basis)
B2.2		Low (guess)
B2.3		Moderate (weak inference)

B2.4		High (strong inference)
B2.5		Very high (explicitly stated)
B2.6		Unknown/Unable to code

B3. Counterbalancing of Change Agents (answer B3.1 to B3.3)

B3.1		By change agent
B3.2		Statistical
B3.3		Other

Not carried out.

B4. Group Equivalence Established (select one of the following)

B4.1		Random assignment
B4.2		Posthoc matched set
B4.3		Statistical matching
B4.4		Post hoc test for group equivalence

Initially matched in terms of grade and all schools being mixed gender and disadvantaged. Post hoc test to show no significant difference in gender, location (urban/rural) or multi-grade.

B5. Equivalent Mortality (answer B5.1 through B5.3)

B5.1		Low Attrition (less than 20% for Post)
B5.2		Low Attrition (less than 30% for follow-up)
B5.3		Intent to intervene analysis carried out

F. Implementation Fidelity

F1. Evidence of Acceptable Adherence (answer F1.1 through F1.3)

F1.1		Ongoing supervision/consultation	
F1.2		Coding intervention sessions/lessons or procedures	
F1.3		Audio/video tape implementation (select F1.3.1 or F1.3.2):	
		F1.3.1	Entire intervention
		F1.3.2	Part of intervention

2-day training workshop. Ongoing supervision through school visits and an interim group meeting.

F2. Manualization (select all that apply)

F2.1		Written material involving a detailed account of the exact procedures and the sequence in which they are to be used
F2.2		Formal training session that includes a detailed account of the exact

		procedures and the sequence in which they are to be used
F2.3		Written material involving an overview of broad principles and a description of the intervention phases
F2.4		Formal or informal training session involving an overview of broad principles and a description of the intervention phases

The content of the training is not explicitly stated, but as it was for 2 days it is likely to have been detailed, as typical training for the programme lasts one day and it is known that the programme has a specified lesson structure.

F3. Adaptation procedures are specified (select one)

	Yes
	No
	Unknown

Rating for Implementation Fidelity (select 0, 1, 2, or 3):

3	
2	
1	
0	

Fidelity high = both type one (86.4%) and type 2 (86.6%) intervention groups reported delivery of the full programme. Structured observations (n=27) were carried out over the course of implementation and found 90.9% of observed activities to be implemented fully, supporting this.

Rating of 2 given, as whether written materials were given on training days is not explicitly stated, so cannot be sure of level of manualisation.

I. Follow up assessment

	Timing of follow up assessment: specify	12 month follow up
	Number of participants included in the follow up assessment: specify	Type 1 intervention: 205 Type 2 intervention: 222 Control: 185
	Consistency of assessment method used: specify	Same measures administered. SDQ measures administered on follow up were subscales that had shown no effects at post intervention to assess whether impact was delayed.

Overall rating for follow up assessment:

3	
2	
1	
0	

Follow up occurred at one additional time point (12 months).

III. Other Descriptive or Supplemental Criteria to Consider

A. External Validity Indicators

A1. Sampling procedures described in detail

	Yes
	No

Specify rationale for selection:

Disadvantaged schools used as it was delivered as part of 'Delivering Equality of Opportunity in Schools' in Ireland and economically disadvantaged are more at risk of social/emotional/behavioural difficulties

Specify rationale for sample size: No rationale given

A1.1 Inclusion/exclusion criteria specified

	Yes
	No

To be included schools had to be mixed gender, have disadvantaged status (from Department of Education and Skills) and classes needed 10+ children. Location in Ireland was also noted.

A1.2 Inclusion/exclusion criteria similar to school practice

	Yes
	No

Intervention would not be withheld based on single sex schooling or small class sizes in typical school practice.

A1.3 Specified criteria related to concern

	Yes
	No

Disadvantaged pupils of relevance to increasing coping strategies, as they are more at risk of negative life events.

A2. Participant Characteristics Specified for Treatment and Control Group

Participants from Treatment Group 1	Grade/age	Gender	Ethnicity/race or Multi-ethnic/ ethnic identity	Primary Language	SES	Family Structure	Locale	Disability	Functional Descriptors
Child/Student Parent/caregiver Teacher School Other	First class (7-8 years)	132 male, 135 female	unknown	unknown	Disadvantaged school, not detailed at individual pupil level	Unknown	Ireland 76 urban, 191 rural	Unknown	Disadvantages schools

Participants from Treatment Group 1	Grade/age	Gender	Ethnicity/race or Multi-ethnic	Primary Language	SES	Family Structure	Locale	Disability	Functional Descriptors
Child/Student Parent/caregiver Teacher School Other	First class (7-8 years)	134 male, 117 female	unknown	unknown	Disadvantaged school, not detailed at individual pupil level	Unknown	Ireland 86 urban, 165 rural	Unknown	Disadvantages schools

Participants from Control Group	Grade/age	Gender	Ethnicity/race or Multi-ethnic	Primary Language	SES	Family Structure	Locale	Disability	Functional Descriptors
Child/Student Parent/caregiver Teacher School Other	First class (7-8 years)	113 male, 99 female	unknown	unknown	Disadvantaged school, not detailed at individual pupil level	unknown	Ireland 239 urban, 491 rural	Unknown	Disadvantages schools

Characteristics of teachers delivering the programme not given (other than 18 for both intervention groups and 16 for the control group). ‘Health promotion specialists’ delivered training and supervision, but no further participant characteristics relating to them were given.

A3. Details are provided regarding variables that:

A3.1 Have differential relevance for intended outcomes

	Yes
	No

Specify: details given on gender for each group

A3.2 Have relevance to inclusion criteria

	Yes
	No

Specify: only delivered in disadvantaged schools, but no measure of individual SES was taken.

A4. Receptivity/acceptance by target participant population (treatment group)

Participants from treatment group	Results (What person reported to have gained from participation in program)	General Rating
Child/Student Parent/caregiver Teacher School Other	Reactions from pupils not stated.	Participants reported benefiting overall from the intervention Participants reported not benefiting overall from the intervention
Child/Student Parent/caregiver Teacher School Other	Reactions from teachers were indicated to have been positive, “these positive results were supported by additional qualitative findings from the teachers’ end of programme review questionnaire and focus group sessions” (p.794), but were not explicitly stated. However references to other papers were given: Clarke & Barry, in press Clarke, 2011; Clarke, O’Sullivan & Barry, 2010.	Participants reported benefiting overall from the intervention Participants reported not benefiting overall from the intervention

Participants from control group	Results (What person reported to have gained from participation in	General Rating
---------------------------------	--	----------------

	program)	
Child/Student Parent/caregiver Teacher School Other	Not stated.	Participants reported benefiting overall from the intervention Participants reported not benefiting overall from the intervention
Child/Student Parent/caregiver Teacher School Other	Not stated	Participants reported benefiting overall from the intervention Participants reported not benefiting overall from the intervention

A5. Generalization of Effects: Not stated.

A5.1 Generalization over time: Not stated

A5.1.1 Evidence is provided regarding the sustainability of outcomes after intervention is terminated

	Yes
	No

Specify: emotional literacy gains persisted at 12-month follow up, with the exception of the subscale on empathy.

A5.1.2 Procedures for maintaining outcomes are specified

	Yes
	No

Specify:

A5.2 Generalization across settings

	Yes
	No

Specify:

A5.2.1 Evidence is provided regarding the extent to which outcomes are manifested in contexts that are different from the intervention context

	Yes
	No

A5.2.2 Documentation of efforts to ensure application of intervention to other settings

	Yes
	No

Specify:

A5.2.3 Impact on implementers or context is sustained

	Yes
	No

Specify: focus of change was the children, not context or implementers

A5.3 Generalization across persons

Evidence is provided regarding the degree to which outcomes are manifested with participants who are different than the original group of participants for with the intervention was evaluated

	Yes
	No

Specify: Although ceiling effects due to having a non-clinical sample is mentioned

B. Length of Intervention (select B1 or B2)

B1.		Unknown/insufficient information provided
B2.		Information provided. If information is provided, specify one of the following:
	B2.1	Weeks
	B2.2	months
	B2.3	Years 1 (academic). Assumed, as stated in introduction that Zippy's Friends should be implemented over one academic year
	B2.4	other

C. Intensity/dosage of Intervention (select C1 or C2)

C1.		Unknown/insufficient information provided
C2.		Information provided. If information is provided, specify one of the following:
	C2.1	length of intervention session 1 hour
	C2.2	frequency of intervention session once a week

This is assumed based on information given in the introduction and the study later specifying high adherence to the programme

D. Dosage Response (select D1 or D2)

D1.		Unknown/insufficient information provided
D2.		Information provided. If information is provided, answer D2.1

	D2.1	Describe positive outcomes associated with higher dosage:
--	------	---

E. Program Implementer (select all that apply)

E1.		Research Staff
E2.		School Specialty Staff
E3.		Teachers
E4.		Educational Assistants
E5.		Parents
E6.		College Students
E7.		Peers
E8.		Other
E9.		Unknown/insufficient information provided

F. Characteristics of the Intervener

F1.		Highly similar to target participants on key variables (e.g., race, gender, SES)
F2.		Somewhat similar to target participants on key variables
F3.		Different from target participants on key variables

Characteristics of teachers implementing the intervention not given

H. Cost Analysis Data (select G1 or G2)

H1.		Unknown/insufficient information provided
H2.		Information provided (if information is provided, answer H2.1)
	H2.1	Estimated Cost of Implementation

I. Training and Support Resources (select all that apply)

I1.		Simple orientation given to change agents
I2.		Training workshops conducted
		#of Workshops provided 2
		Average length of training day long sessions x2
		Who conducted training (select all that apply)
	I2.1	Project Director
	I2.2	Graduate/project assistants
	I2.3	Other (please specify): Health promotion specialists
	I2.4	Unknown
I3.		Ongoing technical support
I4.		Program materials obtained
I5.		Special Facilities
I6.		Other (specify):

J. Feasibility

J1.	Level of difficulty in training intervention agents (select one of the following)	
	J1.1	High

	J1.2		Moderate
	J1.3		Low
	J1.4		Unknown
J2.	Cost to train intervention agents (specify if known): unknown		
J3.	Rating of cost to train intervention agents (select one of the following)		
	J3.1		High
	J3.2		Moderate
	J3.3		Low
	J3.4		Unknown

Appendix F

Weight of Evidence A

Tables 1-4 detail the criteria needed to be met for scoring of the measures, comparison group, fidelity and follow up variables.

This rationale has been derived from the 'II Key Features for Coding Studies and Rating Level of Evidence/ Support; section of the Kratochwill (2003) coding protocol.

Table 1
Measurement (A.1 – A.4 of Kratochwill 2003 Coding Protocol)

Weighting	Description
High (3)	<ul style="list-style-type: none"> • Reliability of .85 or higher (taking into account the population of the sample) • Multiple methods used • Multiple sources used • Validity statistic must be reported • All primary outcomes measures must meet the above criteria
Medium (2)	<ul style="list-style-type: none"> • Reliability of measures used must be .7 or higher • Multiple methods OR multiple sources used. • Validity is not necessary to be included • 75% of primary outcomes measures must meet the above criteria
Low (1)	<ul style="list-style-type: none"> • Reliability of measures of at least .5 • Validity is not necessary to be included • 50% of primary outcomes measures must meet the above criteria
(0)	<ul style="list-style-type: none"> • Reliable measure not used and it is the only method/source of measurement

Table 2
Comparison Group (B1-B5 of Kratochwill 2003 Coding Protocol)

Weighting	Description
High (3)	<ul style="list-style-type: none"> • Active comparison group e.g. alternative intervention or attention placebo • Equivalence of groups • Equivalent mortality and low attrition (including for follow ups where applicable) • Change agents counterbalanced
Medium (2)	<ul style="list-style-type: none"> • ‘No intervention’ comparison group e.g. waitlist or no intervention • Equivalence of groups OR equivalent mortality and low attrition for each group OR counterbalancing of change agents • If no equivalent mortality, analysis must confirm there are no significant differences between groups
Low (1)	<ul style="list-style-type: none"> • Comparison group included in the study • 1 of the following must be present: Equivalence of groups OR equivalent mortality and low attrition for each group OR counterbalancing of change agents. • If no equivalent mortality, analysis must confirm there are no significant differences between groups
No evidence (0)	<ul style="list-style-type: none"> • No effort given to ensure the equivalence of groups

Table 3
Fidelity (F1-F3 of Kratochwill 2003 Coding Protocol)

Weighting	Description
High (3)	Acceptable adherence demonstrated through use of a manual AND two of the following: supervision, coding sessions or recording (video or audio) the session. 'Manual' can be demonstrated either by use of training or written materials given which give exact details of the procedure and sequence for intervention implementation. Procedures for adaptation given.
Medium (2)	Acceptable adherence demonstrated through use of a manual AND one of the following: supervision, coding sessions or recording (video or audio) the session. 'Manual' can be demonstrated either by use of training or written materials given which give broad overview of principles/intervention phases
Low (1)	Acceptable adherence demonstrated through use of a manual or one of the following: supervision, coding sessions or recording (video or audio) the session
No evidence (0)	Nothing done to ensure fidelity or adherence is not acceptable

Table 4
Follow up assessment conducted (sub-section 1 of Kratochwill 2003 Coding Protocol)

Weighting	Description
High (3)	Follow-up assessments carried out over multiple intervals, with all original participants and using similar measures to the original measurement
Medium (2)	At least one follow-up assessments carried out, with the majority of the original sample, and using the similar measures to the original measurement
Low (1)	At least one follow up using some original participants
No evidence (0)	No follow up

Table 5
Statistical Analysis (sub-section C of Kratochwill 2003 Coding Protocol)

Weighting	Description
High (3)	Appropriate statistical analysis conducted AND appropriate units of analysis AND familywise error controlled where applicable AND a sufficiently large N to detect a small effect size.
Medium (2)	Appropriate statistical analysis conducted AND appropriate units of analysis AND familywise error controlled where applicable AND a sufficiently large N to detect a medium effect size
Low (1)	Appropriate statistical analysis conducted AND appropriate units of analysis AND familywise error controlled where applicable.
No evidence (0)	None of the above criteria are met

The protocol in Table 5 has been adapted to take into account the size of the effect the study has the power to detect. Giving more weight to studies that have enough participants to detect a small effect size, as previous literature has found universal social and emotional interventions to result in effect sizes ranging from small to medium.

The overall quality of each study was calculated by adding the scores allocated for the different variables (measures, comparison group, fidelity and follow up) and dividing them by 4. The overall quality categories are identified as follows:

Table 6
Average WoE A score descriptors

Overall quality	Averaged score
High	2.5 or higher
Medium	1.5 to 2.4
Low	1.4 or less

Table 7
WoE A scores for the studies included in the review

Studies	Measures	Comparison Group	Fidelity	Follow up	Statistical Analysis	Average score (WoE A)*
Clarke, Bunting & Barry (2014)	2	2	2	2	2	Medium (2)
Dufour , Denoncourt & Mishara (2011)	1	1	2	0	2	Low (1.2)
Holen , Waaktaar , Lervåg & Ystgaard (2012)	1	2	3	0	3	Medium (1.8)
Mishara & Ystgaard (2006)	1	1	3	0	2	Low (1.4)
Rodker (2013)	1	0	0	0	1	Low (0.4)

*Average WoE A scores have been rounded to one decimal place.

Appendix G

Weight of evidence B: Methodological relevance to the review question

Table 1
Appropriateness of measures

Rating	Description
High rating (3)	Pre, post and follow up measures are taken for both groups.
Medium rating (2)	Pre and post measures are taken for both groups, with post measures being taken immediately after the intervention has finished.
Low rating (1)	Pre and post measures are taken. Post measure may not be taken immediately after the intervention has finished.

Rationale: Having both pre and post measures and a follow up allows for investigation of the durability of the intervention effects, which is of relevance to this review in order to assess the intervention’s effectiveness. Post-tests that occur immediately after the intervention are favourable to those taken after an elapsed time because additional intervention could have occurred within an elapsed time or intervention effects may have dissipated.

Table 2
Sources of measurement

Rating	Description
High rating (3)	Measurement is taken from three or more sources e.g. parents, teachers and students
Medium rating (2)	Measurement is taken from two sources e.g. students and teachers
Low rating (1)	Measurement is taken from one source e.g. student only

Rationale: Kratochwill (2003) advises measuring the primary outcome variable using at least two measures. This is of particular importance for this review, as participants were children aged 5-8 years, so they may have had difficulty accurately self-reporting; additional sources therefore would allow for triangulation of the information obtained. Additionally, teacher and parent reports have often been used and multiple sources allows for differing behaviours across home and school to be incorporated into the measurement. Due to the aim of Zippy’s Friends being to equip children with coping skills which they can use in response to future life event, it would be expected that changes in use of coping skills would be expected to generalise to home. Multiple sources using both parents and teachers would allow an interpretation of whether this is occurring.

Table 3
Comparison group

Rating	Description
High rating (3)	Both active control and waitlist/no intervention control group used
Medium rating (2)	Active control group used
Low rating (1)	No intervention control group used

Rationale: Having an active control group is deemed more beneficial than a no intervention control group because it allows us to establish if Zippy’s Friends is effective in causing change beyond that of another intervention/attention control. This will determine whether there is a benefit of recommending this specific intervention over alternative support. A no intervention control group is required as a minimum, to establish if benefits of the Zippy’s Friends intervention go beyond that of maturation. Having both an active and no intervention control is considered ideal as it allows comparisons across the three groups; this would allow the researcher to note if, in fact, both the intervention and active control groups are both not significantly different from a no intervention control.

Table 4
Summary table of weight of evidence B

Study	Appropriateness of measures	Sources of measurement	Comparison group	Average WoE B**
Clarke, Bunting & Barry (2014)	3	1	3	2.3
Dufour , Denoncourt & Mishara (2011)	2	3	1	2
Holen , Waaktaar , Lervåg & Ystgaard (2012)	2	3	1	2
Mishara & Ystgaard (2006)	2	2	0.5*	1.5
Rodker (2013)	1	3	1	1.6

*Lithuanian sample had control groups, but for the Denmark sample, a previous control group’s data was used due to financial and time constraints on data collection. 0.5 has been allocated for this study to acknowledge the lack of control group for all students within the study. More than ‘0’ was allocated, as a control group were used for over half of participants and the data analysed for different groups separately.

**Average WoE B scores have been rounded to one decimal place.

Appendix H

Weight of Evidence C - Topic relevance to the review question

Table 1
WoE C ratings

Weighting	Description
High (3)	Programme implemented as intended. AND training given for at least one full day prior to implementation AND ongoing support throughout implementation of the programme.
Medium (2)	Programme implemented as intended AND training given for at least one full day prior to implementation AND follow up supervision OR ongoing support throughout implementation of the programme.
Low (1)	Programme implemented with some changes or fidelity not monitored

Rational: Due to the research question considering the effectiveness of the Zippy's Friend's programme, Weight of Evidence C takes into account the fidelity to the intervention, training prior to implementation and ongoing support during implementation. Differences in training may have affected the programme being delivered as intended and may have influenced the confidence and self-efficacy beliefs of teachers implementing the programme. Presence and level of supervision may have affected the ability of those delivering the programme to overcome any obstacles encountered in its delivery. Presence of supervision and lack of prior training both goes against recommendations for using Zippy's Friends, and may have affected it being implemented as intended. Therefore, in order to assess effectiveness of the programme, those that have implemented Zippy's Friends in its intended format, with recommended training and supervision, will be the most useful papers to answer the review question.

Table 2
WoE C Summary table for reviewed studies

Study	WoE C Score	WoE C Descriptor
Clarke, Bunting & Barry (2014)	3	High
Dufour , Denoncourt & Mishara (2011)	2	Medium
Holen , Waaktaar , Lervåg & Ystgaard (2012)	2	Medium
Mishara & Ystgaard (2006)	3	High
Rodker (2013)	1	Low

Appendix I

Weight of Evidence D

$$\frac{\text{WoE A} + \text{WoE B} + \text{WoE C}}{3} = \text{WoE D}$$

3

Table 1
Overall weight of evidence

Studies	WoE A	WoE B	WoE C	WoE D*
Clarke, Bunting & Barry (2014)	Medium (2)	Medium (2.3)	High (3)	Medium (2.4)
Dufour , Denoncourt & Mishara (2011)	Low (1.2)	Medium (2)	Medium (2)	Medium (1.7)
Holen , Waaktaar , Lervåg & Ystgaard (2012)	Medium (1.8)	Medium (2)	Medium (2)	Medium (1.9)
Mishara & Ystgaard (2006)	Low (1.4)	Medium (1.5)	High (3)	Medium (2)
Rodker (2013)	Low (0.4)	Medium (1.6)	Low (1)	Low (1)

*WoE D scores have been rounded to one decimal place.

The following table shows the criteria for the different categorical ratings of WoE D. This criteria ensures that for a study to be deemed 'high' for WoE D, it must have been deemed 'high' in at least two of the other WoE ratings.

Table 2
WoE D categories

Category	Numerical Rating
Low	1.4 or less
Medium	1.5-2.4
High	2.5 or above