

Case study 1: An Evidence-based practice review report.

*Theme: School/Setting Based Interventions for Social, Emotional and
Mental Health.*

The Effectiveness of MindUP

The 'MindUP Program', herein referred to as 'MindUP' is a mindfulness-based intervention (MBI) for schools developed by The Goldie Hawn Foundation (TGHF) (n.d.) to respond to social, emotional and mental health (SEMH) difficulties in children and young people (CYP). This systematic literature review aims to evaluate how effective MindUP is for reducing internalising problems in CYP aged 3 to 14 years old.

The review was undertaken using five online databases: PsycINFO, Science Direct, Scopus, Web of Science and ERIC (EBSCOhost). Five studies were included for the review and evaluated using Gough's (2007) Weight of Evidence Framework and the coding protocol by Gersten et al. (2005). All studies used an experimental group design with outcome measures collected at pre and post-intervention.

Results showed small effects for outcome measures related to internalising problems at post-intervention although the effects were larger than when the culturally adapted version of MindUP was used. The findings are discussed, and limitations and future recommendations are considered.

Introduction

Mental health difficulties have been impacting around 20% of children and young people (CYP) across the globe with 50% of them experiencing an

onset before 14 years old (Department of Health [DH], 2015). CYP's psychosocial development and mental health difficulties need to be addressed before they become severe long-term conditions extending into adulthood (Department for Education [DfE], 2019). SEMH needs have been included as part of the special educational provision within schools (DfE & DH, 2015), and it is the responsibility of professionals working with CYP to regularly track whether children are developing well to meet their developmental milestones within the typical age range in comparison to peers of the same age especially in early years settings – early tracking and identifying needs in the early years setting have been made statutory through the Early Years Foundation Stage framework (DfE, 2021).

Mindfulness and its benefits on SEMH

A growing body of systematic literature reviews show that mindfulness-based interventions (MBIs) within education around the world have made contributions to CYP's social and emotional learning (SEL) with positive effects on SEMH (e.g. Corcoran et al., 2018; Maynard et al., 2017; Weare, 2019). Dunning et al. (2018) conducted a meta-analysis on MBIs that only included Randomised Controlled Trials (RCTs) and found significant effects on the reduction of anxiety and depression after participants went through a course of MBI compared to a control group.

There are many types of MBIs. For example, Kuyken et al. (2013) investigated the effectiveness of the Mindfulness in Schools Program (MiSP) in the United Kingdom (UK) while Sibinga et al. (2016) implemented the Mindfulness-based Stress Reduction (MBSR) program in Baltimore, United States (US). Both studies delivered to secondary school pupils found a

reduction in symptoms of depression and comorbid conditions in the intervention group.

Another intervention called MindUP is now widely included in the primary school curriculum in British Columbia and Newark, New Jersey due to its effectiveness (Sisk, 2018). In light of recent research done to evaluate the benefits of MindUP, it would be valuable to find out its specific impact on SEMH difficulties (Weare, 2019).

What is MindUP?

MindUP for schools was developed by The Goldie Hawn Foundation (TGHF) (n.d.) to respond to the global mental health crisis of childhood anxiety, depression, aggression and suicide (World Health Organization, n.d.). It is a universal and inclusive intervention for CYP 3 to 14 years old which means that MindUP is made available to all CYP including those with special needs or disabilities (DfE & DH, 2015). MindUP aims to improve pro-social behaviours, attentiveness, emotional regulation, reduce aggression, and improve math and literacy skills (Musty, 2015). It hopes to help CYP manage stressful encounters, build resilience and optimism, and cultivate compassion for others (Sisk, 2018; TGHF, n.d.). MindUP was evaluated by the Collaborative for Academic, Social, and Emotional Learning (CASEL) and received the 'SElect' status (CASEL, 2013). This means that MindUP has been recognized as a programme that is well-designed to promote CYP's social and emotional competence at the universal level which includes providing quality training to professionals who will implement the programme (CASEL, 2013). MindUP has shown evidence-based effectiveness on CYP's

SEL through evaluation of the intervention using a comparison group and pre-post measures (CASEL, 2013).

Teachers implementing MindUP will typically receive initial face-to-face group training lasting 4 to 6 hours from a certified MindUP consultant. The school will then implement the curriculum supported by the consultant via video calls. Usually, in the middle of the academic year, a follow-up face-to-face group training will take place where the consultant also conducts classroom observations and tailored mentoring for teachers. Towards the end of the academic year, the consultant and school leadership team will meet to evaluate progress and plan next steps.

How is MindUP delivered?

MindUP is a manualised curriculum delivered by trained teachers in classrooms. During each lesson, pupils will learn mindfulness techniques through activities tailored to the objectives of each lesson. Teachers will facilitate the generalisation of skills from lessons throughout the entire day so that pupils will internalise the skills and play a part in creating a positive classroom environment (Maloney et al., 2016).

Main features of MindUP

MindUP consists of 15 detailed lesson plans to be taught weekly, with 'Core Practices' – involving focused breathing – three times a day after completion of Unit 1 (Maloney et al., 2016). The duration of each lesson and 'Core Practices' are tailored according to age with shorter sessions for younger children and longer sessions for older children. Pupils learn about how their brain works, understand their own thoughts, gain awareness of

their own internal experiences, and actively consider their impact on the world (Maloney et al., 2016). Table 1 shows the curriculum.

Table 1
MindUP Curriculum

Unit	Lesson	Topic
1: Getting Focused	1	How Our Brains Work
	2	Mindful Awareness
	3	Focused Awareness: The Core Practice
2: Sharpening Your Senses	4	Mindful Listening
	5	Mindful Seeing
	6	Mindful Smelling
	7	Mindful Tasting
	8	Mindful Movement I
	9	Mindful Movement II
3: It's All about Attitude	10	Perspective Taking
	11	Choosing Optimism
	12	Appreciating Happy Memories
4: Taking Action Mindfully	13	Expressing Gratitude
	14	Performing Acts of Kindness
	15	Taking Mindful Action in the World

Psychological underpinning for MindUP

The MindUP curriculum is grounded on four theories: Firstly, contemplative science involves understanding the effects of mindfulness on the body and the mind across the lifespan; this theory is applied to creating

age-appropriate interventions to support child development and teachers' wellbeing in schools (Roeser & Zelazo, 2012).

Secondly, high-quality education should include teaching on SEL where CYP learn how to interact with each other in socially respectable ways and are provided with opportunities to practice these skills in a safe place such as school settings before they enter the workforce after completing their education (Greenberg et al., 2003). Acquiring skills like self-confidence, interpersonal problem-solving and communication greatly curbs SEMH difficulties in CYP (Durlak & Wells, 1997).

Thirdly, theories from positive psychology suggests that happiness and positivity are governed by three main factors: First, the type of gene expression for happiness plays a part in influencing the extend of happiness a person can feel; second, circumstances that facilitate the feelings of happiness; third, participating in activities that are known to bring happiness (Lyubomirsky et al., 2005). Recognizing these factors enables us to plan interventions that can create an environment that facilitates feelings of happiness (Lyubomirsky & Layous, 2013). This encourages participants to engage in activities that will increase their positive affect which can increase overall mental wellbeing (Lyubomirsky & Layous, 2013).

Lastly, cognitive developmental neuroscience informs how CYP's brains develop in relation to their environment. For example, what a person experiences affects gene expression or a person's genes could also affect how they would experience their environment (Diamond, 2009; Zelazo & Lyons, 2012). Understanding these links between the mind and the body allows participants to realize that they have the ability to control their

thoughts which indirectly impact their display of different behaviours. Thus, such strategies allow participants who have received MindUP to increase their own locus of control and experience fewer negative affect.

Aims and review question

With substantial evidence of MBIs yielding positive effects on internalising difficulties and a strong psychological basis, the main aim of this literature review is to find out how effective MindUP can be used as a universal school-based intervention to address internalising difficulties in CYP. This will inform policies and practices that are intrinsic to the work of Educational Psychologists (EPs) by providing evidence-based informed intervention to schools that can benefit from MindUP. Specifically, the reviewer aims to address the following research question: “How effective is MindUP for reducing internalising problems in children ages 3 to 14 years old?”

Critical Review of the Evidence Base

Literature search

On 21st January 2021, a literature search was conducted on PsycINFO, Science Direct, Scopus, Web of Science and ERIC (EBSCOhost). The following search terms in Table 2 were used to identify potential studies.

Table 2

Search Terms

Databases searched	Search Terms
PsycINFO	MindUP or MindUP* or "Mind UP" or "Mind UP*"
ERIC (EBSCOhost)	(searched using 'full text' option)
Scopus	AND
Web of Science	

Science Direct	School or Kindergarten or nursery or pre*school* or student or children or preschooler or pre*schooler or pupil*
----------------	--

Figure 1 below shows how the full literature search process was carried out. 271 studies were yielded. 11 were duplicates, 12 were not written in English and 75 were not peer-reviewed. The reviewer screened through titles and abstracts at the same time for all 173 studies with reference to the inclusion and exclusion criteria (see Table 3) to ensure that the studies were not simply excluded if they did not mention 'MindUP' in their title. 167 studies were excluded as they were not delivering MindUP. A full text screening was undertaken for the remainder 6 studies. 3 studies were eventually excluded because MindUP only informed parts of the mindfulness intervention delivered (see Appendix A for full rationale). An ancestral search was conducted and 2 extra studies were found to be delivering MindUP resulting in the final 5 studies to be reviewed (Table 4).

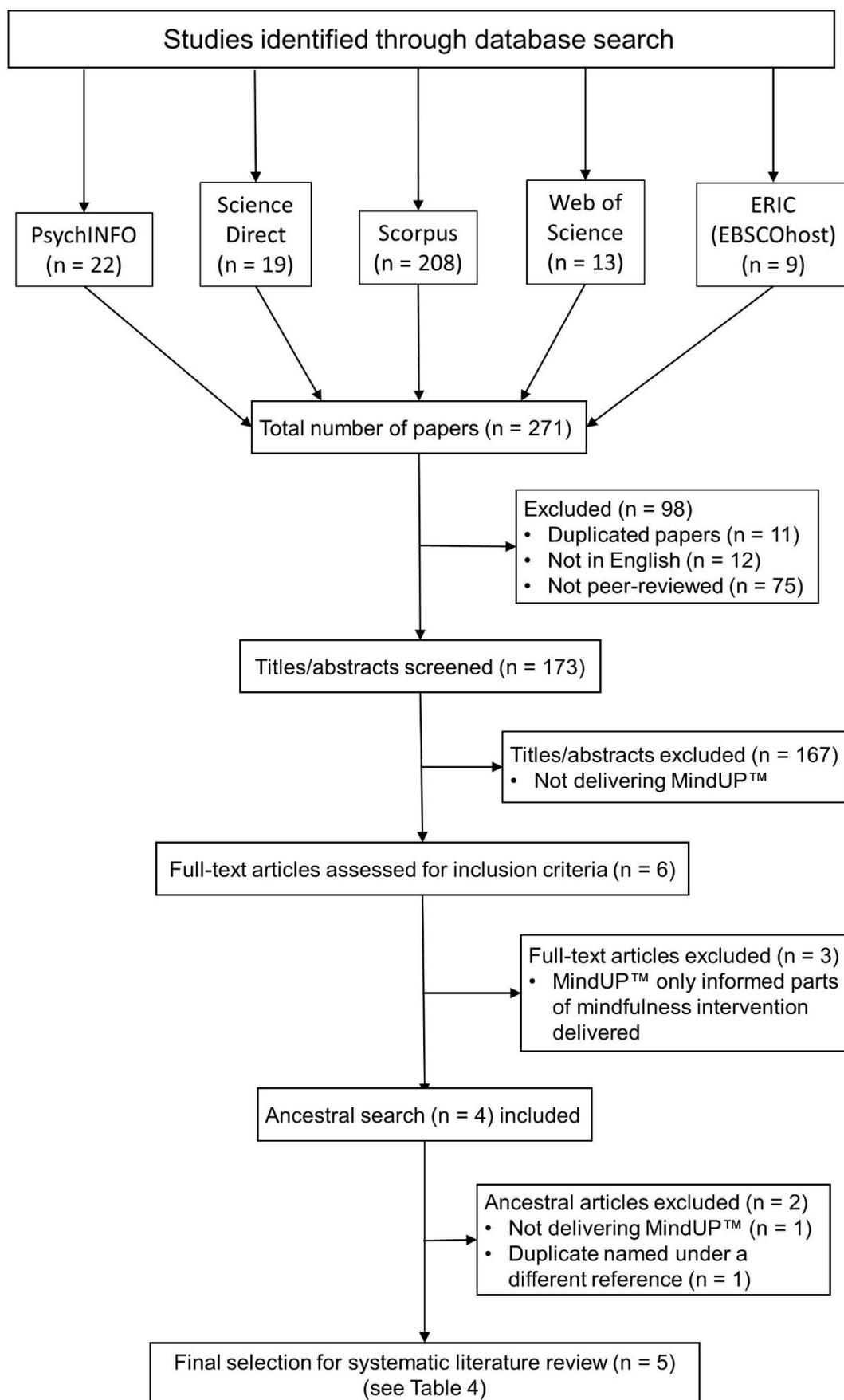


Figure 1. Literature search process

Table 3

Inclusion and Exclusion Criteria

Criterion	Inclusion criteria	Exclusion criteria	Justification
1 Language of publication	Studies published in English.	Studies not published in English.	To ensure that the reviewer is able to critically appraise the studies in the first language.
2 Type of publication	Studies published in a peer-reviewed journal.	Studies not published in a peer-reviewed journal.	Studies from peer-reviewed journals are more likely to have scientific rigour and high quality of research design.
3 Type of Intervention	Study must have delivered MindUP in a school setting.	Study did not deliver MindUP or was not in a school setting	To be able to critically evaluate the effectiveness of MindUP as a school-based intervention for reducing internalising problems.
4 Research design and methodology	Study must use an experimental research design with empirical data collected pre- and post-intervention. Outcome measures must include evaluation of internalising problems related to social and emotional mental wellbeing.	An experimental research design was not used and empirical data was not collected at pre- and post-intervention. Outcome measures did not include evaluation of internalising problems related to social and emotional mental wellbeing.	To be able to review original data. Quantitative data is needed to critically evaluate MindUP for its effectiveness by calculating effect sizes across all studies for comparison. To evaluate the effectiveness of MindUP for reducing internalising problems.
5 Participants	Participants are students attending Pre-Kindergarten to	Participants are outside the age range of 3 to 14 years old.	MindUP was designed to be delivered to students in this age

	8 th Grade (i.e. 3 to 14 years old).		range within schools.
6	Sample size More than 1 participant.	Single case studies involving 1 participant.	To ensure that results from this review are reliable.

Table 4

Studies Included in this Systematic Literature Review

Number	Reference
1	Crooks, C. V., Bax, K., Delaney, A., Kim, H., & Shokoohi, M. (2020). Impact of mindUP among young children: Improvements in behavioral problems, adaptive Skills, and executive functioning. <i>Mindfulness</i> , 11(10), 2433-2444. https://doi.org/10.1007/s12671-020-01460-0
2	de Carvalho, J. S., Pinto, A. M., & Marôco, J. (2017). Results of a mindfulness-based social-emotional learning program on portuguese elementary students and teachers: a quasi-experimental study. <i>Mindfulness</i> , 8(2), 337-350. https://doi.org/10.1007/s12671-016-0603-z
3	Matsuba, M. K., Schonert-Reichl, K. A., McElroy, T., & Katahoire, A. (2020). Effectiveness of a SEL/mindfulness program on Northern Ugandan children. <i>International Journal of School & Educational Psychology</i> , 1-19. https://doi.org/10.1080/21683603.2020.1760977
4	Schonert-Reichl, K. A., Oberle, E., Lawlor, M. S., Abbott, D., Thomson, K., Oberlander, T. F., & Diamond, A. (2015). Enhancing cognitive and social-emotional development through a simple-to-administer mindfulness-based school program for elementary school children: A randomised controlled trial. <i>Developmental psychology</i> , 51(1), 52. https://doi.org/10.1037/a0038454
5	Thierry, K. L., Bryant, H. L., Nobles, S. S., & Norris, K. S. (2016). Two-year impact of a mindfulness-based program on preschoolers' self-regulation and academic performance. <i>Early Education and Development</i> , 27(6), 805-821. https://doi.org/10.1080/10409289.2016.1141616

Weight of Evidence (WoE)

The reviewer used Gough's (2007) Weight of Evidence (WoE) framework to critically appraise the quality and relevance of all five studies included in this review. WoE A appraises the general methodological quality of a study against other similar type of studies (see Appendix B). To do this, the coding protocol by Gersten et al. (2005) was used with all five studies to evaluate the description of participants, procedure for implementing the intervention, outcome measures used and how the data collected was analysed (Appendix C and D). This protocol was chosen as it is deemed appropriate for use with RCTs and quasi-experimental designs were carried out in schools (Gersten et al., 2005).

Next, each study was appraised for the methodological relevance (WoE B) by evaluating whether the type of evidence and design are appropriate for answering the review question (Appendix E). Subsequently, each study was appraised for its suitability and characteristics of the research evidence (WoE C) in answering the review question (Appendix F). Coding protocols were developed by the reviewer to evaluate WoE B and C respectively. The WoE D rating is calculated by averaging the ratings of WoE A, B and C to give a rating for the overall quality of each study and appraises its relevance to this review question. Table 5 shows a summary of the WoE ratings for each study.

Table 5

Summary of Weight of Evidence (WoE) ratings

Studies	WoE A	WoE B	WoE C	WoE D
Crooks et al. (2020)	3 (High)	2 (Medium)	2 (Medium)	2.33 (Medium)
de Carvalho et al. (2017)	1 (Low)	1.67 (Medium)	3 (High)	1.89 (Medium)
Matsuba et al. (2020)	1 (Low)	1.67 (Medium)	2.5 (High)	1.72 (Medium)
Schonert-Reichl et al. (2015)	3 (High)	2 (Medium)	2.5 (High)	2.5 (High)
Thierry et al. (2016)	3 (High)	2.33 (Medium)	1.5 (Medium)	2.28 (Medium)

Note. 1.4 or less = Low, 1.5 to 2.4 = Medium, 2.5 to 3 = High.

Mapping the field

All five studies included ‘intervention versus comparison group’ research designs to evaluate the effectiveness of MindUP. Details regarding participants, procedures and outcomes are summarised in Table 6 below.

Table 6

Mapping the Field: Overview of the Five Included Studies

Study	Child participant characteristics	Research design	Intervention delivery/training	Measures relevant to this literature review	Significant outcomes
Crooks et al. (2020)	<p><u>Country</u> Southwestern Ontario, Canada. Kindergarten schools were located in a Catholic school district.</p> <p><u>Intervention group (n = 261)</u> 10 schools, 23 classrooms. Number of pupils in each class (22.42). Females (52.11%). Mean age of children (4.27 years, SD = 0.61). White ethnic background (67.06%). High Social Risk Index (SRI) (92.72%). Teaching experience of teachers (17.02 years).</p>	<p>Quasi-experimental design. Pre- and post-tests.</p> <p>Intervention group versus comparison group.</p> <p>The school district board selected schools that had high SRI scores – SRIs indicate the presence of disadvantage characteristics such as lower income, unemployment, children with lone-parent etc.</p>	<p>MindUP program:</p> <ul style="list-style-type: none"> 15 weekly lessons, 10 to 15 minutes each. Core mindfulness activity 3 times a day for 1 or more minutes each. <p>Teachers were trained by TGHF on the implementation of MindUP</p> <ul style="list-style-type: none"> First full day training at the start of the academic year Second full day training at the middle of the academic year <p>Implementation start date began whenever teachers completed the baseline measures.</p>	<p>Teacher reports collected at two time points:</p> <ul style="list-style-type: none"> Pre-intervention between September to December 2017 Post-intervention between May to June 2018 <p>Behavior Assessment System for Children, Third Edition, Teacher Rating Scales (BASC-3 TRS; Reynolds & Kamphaus, 2015)</p> <ul style="list-style-type: none"> Behavioral Symptom Index (BSI) Internalising Problems composite scale Externalising Problems composite scale Adaptive Skills composite scale 	<p>Overall, MindUP had the biggest impact on reducing deficits in executive functioning.</p> <p>A multilevel linear regression model was used to analyse the data. Outcomes from the adjusted model reported below.</p> <p><u>BSI</u> There was a significant reduction in scores in the intervention group compared to the comparison group (adjusted $\beta = -2.47$, SE = 0.5, $p < 0.001$). They exhibited less behavioural symptoms post-intervention.</p> <p><u>Internalising Problems composite scale</u> There was a reduction in scores in the intervention group compared to the comparison group (adjusted $\beta = -2.73$, SE = 0.7, $p < 0.001$). They had fewer internalising problems over time while those in the comparison group had worsen post-intervention.</p>

Study	Child participant characteristics	Research design	Intervention delivery/training	Measures relevant to this literature review	Significant outcomes
	<p><u>Comparison group (n = 323)</u> 7 schools, 19 classrooms. Number of pupils in each class (23.51). Females (51.08%). Mean age of children (4.45 years, SD = 0.55). White ethnic background (72.58%). High SRI (36.84%). Teaching experience of teachers (17.77 years).</p>		<p>Teachers filled out lesson tracking sheets indicating the following to assess implementation fidelity:</p> <ul style="list-style-type: none"> • Time spent delivering the intervention • Curricular components completed • Intervention completion date 	<p>Behavior Rating Inventory of Executive Function-Preschool Version (BRIEF-P; Gioia et al., 1996) or Child Version (BRIEF-2; Gioia et al., 2000)</p> <ul style="list-style-type: none"> • Global Executive Composite (GEC) 	<p><u>Externalising Problems composite scale</u> There was a higher reduction in scores in the intervention group compared to the comparison group (adjusted $\beta = -2.05$, SE = 0.5, $p < 0.001$). They exhibited fewer externalising problems post-intervention.</p> <p><u>Adaptive Skills composite scale</u> There was a significant increase in scores in the intervention group compared to the comparison group (adjusted $\beta = 3.45$, SE = 0.6, $p < 0.001$). Although both groups improved significantly, the intervention group improved at a faster rate.</p> <p><u>GEC</u> There was a significant reduction in scores in the intervention group compared to the comparison group (adjusted $\beta = -4.51$, SE = 0.7, $p < 0.001$). Although both groups improved significantly, the intervention group improved at a faster rate.</p>

Study	Child participant characteristics	Research design	Intervention delivery/training	Measures relevant to this literature review	Significant outcomes
de Carvalho et al. (2017)	<p><u>Country</u> Lisbon, Portugal. Elementary schools were located in three municipalities in the district of Lisbon. 95% of participants were Portuguese.</p> <p><u>Intervention group (n = 223)</u> Mean age of children (8.5 years, SD = 0.97). 98 girls and 122 boys. 167 pupils in 3rd grade and 56 in 4th grade.</p> <p><u>Comparison group (n = 231)</u> Mean age of children (8.5 years, SD = 1.04). 115 girls and 112 boys. 131 pupils in 3rd</p>	<p>Quasi-experimental design. Pre- and post-tests.</p> <p>Intervention group versus comparison group.</p> <p>Convenience sampling was used to select 12 schools from 4 clusters with middle socioeconomic level. Participants were then randomly distributed to ensure equal group sizes and socio-economic status.</p>	<p>MindUP program:</p> <ul style="list-style-type: none"> 15 weekly lessons, 44 to 60 minutes each. Core mindfulness activity 3 times a day for 3 minutes each. <p>Teachers were trained by an expert on the implementation of MindUP for a total of 50 hrs over 6 sessions. This was 12.5 hrs longer than the training offered by TGHF so that teachers had enough time to get to grips with the content.</p> <p>The teachers were divided into 3 groups for training.</p> <ul style="list-style-type: none"> First training in September/ October 2012 for 25 hrs Second training in January/ February 2013 for 25hrs <p>Implementation start date began after teachers completed the second training course.</p>	<p>Child measures were administered in a group setting during a class period by a member of the research team at two time points:</p> <ul style="list-style-type: none"> Pre-intervention Post-intervention at the end of the academic year. <p>Modified Positive and Negative Affect Schedule for Children (PANAS-C; Laurent et al., 1999).</p> <ul style="list-style-type: none"> Positive affect Negative affect <p>Modified Emotional regulation questionnaire for children and adolescents (ERQ-CA; Gullone & Taffe, 2012)</p> <ul style="list-style-type: none"> Cognitive reappraisal Expressive suppression <p>Modified Self-compassion scale for children (SCS-C; Neff, 2011 unpublished)</p> <ul style="list-style-type: none"> Self-kindness 	<p>Overall, the Cohen's U3 index showed that MindUP provided a gain in positive affect by 8%, a gain in common humanity by 7%, and a reduction in negative affect and suppression by 8%.</p> <p>To show the direction of change, pre- and post-intervention means (M) and standard deviations (SD) were recorded and reported in this format M(SD). The ANCOVA results were also reported. Positive affect increased from 4.06(0.91) to 4.16(0.77) in the intervention group but decreased from 4.09(0.75) to 3.99(0.83) in the comparison group. There is a significant improvement in the intervention group compared to the comparison group for positive affect [F(1, 444) = 5.263, p = 0.022, d = 0.21].</p> <p>Negative affect decreased from 1.97(0.83) to 1.85(0.75) in the intervention group but increased from 1.89(0.76) to 1.93(0.76) in the comparison group. There is a significant decrease in the intervention group compared to the comparison group for negative</p>

Study	Child participant characteristics	Research design	Intervention delivery/training	Measures relevant to this literature review	Significant outcomes
	grade and 100 in 4th grade.		Materials needed for the implementation of MindUP were translated from English into Portuguese by teachers and a linguist.	<ul style="list-style-type: none"> • Common humanity • Self-judgment • Mindfulness Modified Mindful Attention Awareness Scale adapted for children (MAAS-C; Lawlor et al., 2014)	affect [$F(1, 443) = 4.339, p = 0.037, d = 0.20$]. Expressive suppression decreased from 3.00(1.21) to 2.62(1.09) in the intervention group, and from 2.81(1.09) to 2.69(1.17) in the comparison group. There is a significant decrease in the intervention group compared to the comparison group for expressive suppression [$F(1, 435) = 4.288, p = 0.039, d = 0.20$]. Common humanity increased from 2.78(1.04) to 3.08(1.11) in the intervention group and decreased from 3.13(1.90) to 3.12(1.09) in the comparison group. There is a quasi-significant improvement in the intervention group compared to the comparison group for common humanity [$F(1, 435) = 3.317, p = 0.069, d = 0.25$].
Matsuba et al. (2020)	Country Gulu district, Northern Uganda. Two small private primary schools.	Quasi-experimental design. Pre- and post-tests. Intervention group versus comparison group.	Culturally adapted MindUP program: <ul style="list-style-type: none"> • 15 lessons over 2 terms • Core mindfulness activity 3 times a day for 3 minutes each. It 	Self-reported outcome measures were presented orally and in written form to the children at two time points. English, together with Luo translation, were used to support comprehension	Overall, there were significant differences between groups where the comparison group who did not receive the MindUP intervention had significant increases in Anger, Loneliness, Perceived hostility and Perceived rejection [$F(9, 63) = 2.73, p=0.014, d = 1.12$].

Study	Child participant characteristics	Research design	Intervention delivery/training	Measures relevant to this literature review	Significant outcomes
	<p>Study 2: <u>Intervention group</u> (n = 46) Females (65%). Mean age of children (13 years, SD = 1.4 years). Children from Grades 5 and 6.</p> <p><u>Comparison group</u> (n = 36) Females (53%). Mean age of children (12.3 years, SD = 1.6 years). Children from Grades 5 and 6.</p>	<p>Schools expressed interest and willingness to implement the MindUP program</p> <p>One school was randomly selected, and a second school served as a waitlist comparison group.</p>	<p>took place on average 3 days a week for Grade 5 classes and 5 days a week for Grade 6 classes.</p> <ul style="list-style-type: none"> Materials used were adapted to the cultural context with culturally available resources. <p>Two training sessions were provided for the teachers through the adapted MindUP program by a trained local Acholi researcher and 6 Acholi teachers from Study 1.</p> <ul style="list-style-type: none"> First training occurred before the start of the academic year lasting 4 days. Between Terms 2 and 3, a 3 day review workshop was carried out. <p>Implementation start date began after the child self-report measures were</p>	<p>(Translations were done by 4 Acholi research assistants and achieved by group consensus).</p> <ul style="list-style-type: none"> Pre-intervention measures collected within the first 3 weeks of the academic year Post-intervention at the end of the academic year. <p>Outcome measures came from the NIH Toolbox for 8 to 12 year olds (Gershon et al., 2013).</p> <ul style="list-style-type: none"> Anger Fear Sadness Loneliness Emotional support Friendship Perceived hostility Perceived rejection <p>Empathic concern was measured using the Modified Interpersonal Reactivity Index (IRI; Litvack-Miller et al.,</p>	<p>To show the direction of change, pre- and post-intervention means (M) and standard deviations (SD) were recorded and reported in this format M(SD). The ANOVA results were also reported.</p> <p>There was a significant increase in Anger in the comparison group 0.43(0.82) than in the intervention group 0.09(0.76) [F(9, 63) = 5.19, p = 0.032, d = 0.55].</p> <p>There was a significant increase in Loneliness in the comparison group 0.69(1.06) than in the intervention group 0.37(0.74) [F(9, 63) = 4.26, p = 0.047, d = 0.51].</p> <p>There was a significant increase in Perceived hostility in the comparison group 0.90(1.06) than in the intervention group 0.13(1.14) [F(9, 63) = 7.11, p = 0.012, d = 0.63].</p> <p>There was a significant increase in Perceived rejection in the comparison group 0.63(1.08) than in the intervention group - 0.16(0.91) [F(9, 63) = 11.99, p = 0.001, d = 0.81].</p>

Study	Child participant characteristics	Research design	Intervention delivery/training	Measures relevant to this literature review	Significant outcomes
Schonert-Reichl et al. (2015)	<p><u>Country</u> Western Canada. Large public school district in a suburban community. Four elementary schools with similar socio-economic status, racial and ethnic diversity, and achievement level were identified.</p> <p><u>Intervention group (n = 48)</u> Females (46%). Mean age of children (10.16 years, SD = 0.52 years). Children from Grades 4 and 5. First language English (63%), East Asian (27%), other (10%). Living with two parents (77%). Living with single parent (10%). Joint physical custody arrangement (10%).</p>	<p>Randomised controlled trial. Pre- and post-tests.</p> <p>To prevent diffusion effects, only one classroom in each of the four schools were selected for participation.</p> <p>When teachers consented to participation, they knew they had 50% chance of being randomised into the comparison group.</p> <p>Two classes from two schools were selected for the intervention group.</p>	<p>completed during the first 3 weeks of school.</p> <p>The first author was contacted for the following information:</p> <p>Altered MindUP program:</p> <ul style="list-style-type: none"> • 12 weekly lessons, 40 to 50 minutes each. • There were fewer lessons about the brain compared to the 15-lesson curriculum by TGHF. • Core mindfulness activity 3 times a day for 3 minutes each. <p>Teachers were trained by a certified MindUP trainer</p> <ul style="list-style-type: none"> • First full day training at the start of the academic year • Booster session at the middle of the academic year <p>Intervention began after the baseline data were collected and randomization was allocated.</p>	<p>1997, original measure from; Davis, 1983).</p> <p>Child self-report measures were administered in a group setting during a class period by a member of the research team at two time points. Research assistants were blinded to the study conditions. Each item was read aloud to pupils to control for differences in reading abilities.</p> <ul style="list-style-type: none"> • Pre-intervention • Post-intervention at the end of the academic year. <p>Empathy and perspective-taking measured by the Modified Interpersonal Reactivity Index (IRI; Davis, 1983; Schonert et al., 2012).</p> <ul style="list-style-type: none"> • Perspective-taking • Fantasy • Empathic concern • Personal distress 	<p>Overall, the Cohen's U3 index showed that MindUP provided a gain in 20% of self-reported emotional wellbeing and prosocial behaviours.</p> <p>To show the direction of change, difference means (M) and standard deviations (SD) at post-test were recorded and reported in this format M(SD). The ANCOVA results were also reported.</p> <p>Empathy increased in the intervention group 0.27(1.07) but decreased in the control group - 0.19(1.05). There is a significant improvement in the intervention group compared to the control group for empathy [F(1, 97) = 4.42, p = 0.03, d = 0.42].</p> <p>Perspective-taking improved in the intervention group 0.25(0.89) but decreased in the control group - 0.14(1.02). There is a significant improvement in the intervention group compared to the control group for perspective-taking [F(1, 97) = 4.17, p = 0.04, d = 0.40].</p>

Study	Child participant characteristics	Research design	Intervention delivery/training	Measures relevant to this literature review	Significant outcomes
	<p><u>Control group</u> (n = 51) Females (42%). Mean age of children (10.31 years, SD = 0.52 years). Children from Grades 4 and 5. First language English (68%), East Asian (22%), other (10%). Living with two parents (89%). Living with single parent (7%). Joint physical custody arrangement (4%).</p>		<p>Teachers filled out lesson tracking sheets indicating the following to assess implementation fidelity:</p> <ul style="list-style-type: none"> • Whether or not they delivered the lesson • Track and record core practices delivered 	<p>Optimism and Emotional control measured by a subscale from the Resiliency Inventory (RI; Noam & Goldstein, 1998; Song, 2003).</p> <p>School self-concept measured by Marsh's Self-Description Questionnaire (SDQ; Marsh et al., 1984)</p> <p>Depressive symptoms measured by Seattle Personality Questionnaire for Children (SPQC; Kusché et al., 1988).</p> <p>Mindfulness measured by Mindful Attention Awareness Scale adapted for children (MAAS-C; Lawlor et al., 2014).</p> <p>Social responsibility measured by subscale of the Social Goals Questionnaire (Wentzel, 1993).</p>	<p>Optimism increased in the intervention group 0.22(0.82) but decreased in the control group -0.17(0.81). There is a significant improvement in the intervention group compared to the control group for optimism [F(1, 97) = 5.40, p = 0.02, d = 0.48].</p> <p>Emotional control improved in the intervention group 0.31(0.85) but decreased in the control group -0.21(0.91). There is a significant improvement in the intervention group compared to the control group for emotional control [F(1, 97) = 8.78, p = 0.004, d = 0.59].</p> <p>School self-concept improved in the intervention group 0.23(0.87) but decreased in the control group -0.17(0.78). There is a significant improvement in the intervention group compared to the control group for school self-concept [F(1, 97) = 5.60, p = 0.02, d = 0.50].</p> <p>Mindfulness improved in the intervention group 0.34(1.22) but decreased in the control group -0.30(1.10). There is a significant improvement in the intervention group compared to the control</p>

Study	Child participant characteristics	Research design	Intervention delivery/training	Measures relevant to this literature review	Significant outcomes
					<p>group for mindfulness [$F(1, 97) = 7.94, p = 0.006, d = 0.55$].</p> <p>Depressive symptoms decreased in the intervention group $-0.19(0.72)$, and increased in the control group $0.10(0.55)$. There is a significant decrease in the intervention group compared to the control group for expressive suppression [$F(1, 97) = 4.14, p = 0.04, d = -0.45$].</p>
Thierry et al. (2016)	<p><u>Country</u> Southwestern part of the United States of America. Urban elementary school in a large city.</p> <p><u>Intervention group (n = 23)</u> Children experienced the intervention throughout pre-kindergarten and kindergarten years. Females (52%). Mean age of children (4.56 years, SD = 0.33 years).</p>	<p>Quasi-experimental design. Pre- and post-tests.</p> <p>Intervention group was one cohort of pupils versus comparison group which was the previous-year cohort of pupils at the same school.</p>	<p>MindUP program:</p> <ul style="list-style-type: none"> 15 lessons, 20 to 30 minutes each. Teachers spent 2 to 3 weeks covering each lesson over the course of the academic year. Core mindfulness activity 3 times a day. In the first few weeks, the activity lasted 30 seconds; it extended to 60 seconds when children gained experience. 	<p>Teacher and parent reports collected at two time points:</p> <ul style="list-style-type: none"> Beginning and end of the prekindergarten academic year <p>Behavior Rating Inventory of Executive Function-Preschool Version (BRIEF-P; Gioia et al., 2002)</p> <ul style="list-style-type: none"> Inhibit Shift Emotional control Working memory Plan/organise 	<p>Overall, MindUP had the biggest impact on improving working memory and planning/organization in the intervention group. This shows improvement in the cognitive components of self-regulation. There were no significant effects in the behavioural forms of self-regulation which includes the inhibit, shift and emotional control scales.</p> <p>To show the direction of change, difference means (M) and standard deviations (SD) at post-test were recorded and reported in this format M(SD). The ANOVA results were also reported.</p>

Study	Child participant characteristics	Research design	Intervention delivery/training	Measures relevant to this literature review	Significant outcomes
	<p>Hispanic (87%), African American (9%), White (4%). First language Spanish (52%), English (44%), English/Spanish (4%).</p> <p><u>Control group (n = 24)</u> Children from the previous cohort who did not experience the intervention throughout pre-kindergarten and kindergarten years. Females (46%). Mean age of children (4.54 years, SD = 0.27 years). Hispanic (84%), African American (8%), White (8%). First language Spanish (50%), English (42%), English/Spanish (8%).</p>		<p>Intervention group received the curriculum in their prekindergarten year and continued the core practices in their kindergarten year.</p> <p>Teachers (n = 4) were trained by a certified MindUP trainer</p> <ul style="list-style-type: none"> • One full day training at the start of the academic year <p>Teachers filled out surveys indicating the following to assess implementation fidelity:</p> <ul style="list-style-type: none"> • Whether or not they delivered the lesson • Rating the level of student engagement 		<p>There was a significant improvement in working memory in the intervention group 48.04(5.24) compared to the comparison group 56.58(10.80) [F(1, 45) = 11.73, p < .01, d = -1.02].</p> <p>There was a significant improvement in planning/organization in the intervention group 44.17(5.36) compared to the comparison group 55.42(12.65) [F(1, 45) = 15.49, p < .01, d = -1.17].</p>

Participants

Schonert-Reichl et al. (2015) conducted random sampling which gave it a high rating on WoE B. After convenience sampling, Matsuba et al. (2020) managed to randomly select the school for the intervention while de Carvalho et al. (2017) distributed selected participants randomly to the intervention group which gave them a medium WoE B rating. The remaining two studies selected participants based on convenience sampling only and had a low WoE B rating.

Altogether, there are 1219 participants. The sample size varied from 47 to 584. 51% of all the participants were females. This shows even gender distribution, and attrition rates were negligible in Matsuba et al. (2020) and Thierry et al. (2016) so they were not penalised in WoE A. All studies evaluated MindUP on different age groups: ages 3 to 5 years in Crooks et al. (2020), 4 to 6 years in Thierry et al. (2016), 7 to 9 years in de Carvalho et al. (2017), 9 to 11 years in Schonert-Reichl et al. (2015) and 12 to 14 years in Matsuba et al. (2020). The type and location of schools varied from private to public schools, and located in the city (de Carvalho et al., 2017; Matsuba et al., 2020; Thierry et al., 2016) versus the suburbs (Schonert-Reichl et al., 2015). Crooks et al. (2020) was allocated schools that had high Social Risk Index (SRIs) indicating the presence of disadvantage characteristics such as lower income, unemployment and children living with a lone-parent

The studies took place in Canada, Portugal, Uganda, and the US where educational systems are likely to be different. As the reviewer is unfamiliar with the differences, this was not part of the WoE appraisal framework. Considering the full age range to be 3 to 14 years old, the

variation in demographics, and the different types of education systems, generalisability of the findings from this review could be a strength.

Research Design

All studies conducted an experimental study with an intervention and comparison group with pre and post-testing done. None of the studies collected follow-up measures to track long-term effects. Schonert-Reichl et al. (2015) conducted an RCT meeting all the essential quality indicators (Gersten et al., 2005) and was rated high on WoE A. The remaining studies used a quasi-experimental design but this could contribute to ecological validity. Although Crooks et al. (2020) and Thierry et al. (2016) used convenience sampling, they both scored high on WoE A for overall methodological quality.

Matsuba et al. (2020) did not describe the comparison condition (rated low on WoE B) nor the characteristic of the teachers who implemented the intervention. de Carvalho et al. (2017) also did not describe the comparison condition (rated low on WoE B) nor checked the fidelity of implementation. Both were given low ratings on WoE A because the missing information was important for replicability of the study.

There were variations across the described comparison group condition for the three studies that had provided information. Thierry et al. (2016) was rated high on WoE B for clearly mentioning what the intervention group had received that the comparison group did not. Therefore, both Crooks et al. (2020) and Schonert-Reichl et al. (2015) were rated medium on WoE B for only describing the comparison conditions because it is important

to clearly identify the independent variable of interest and to be aware of possible confounding variables that could influence the outcome.

Intervention

Comparing the delivery of intervention, Crooks et al. (2020) and Thierry et al. (2016) delivered the MindUP curriculum as intended and were given high WoE B ratings. While de Carvalho et al. (2017) and Matsuba et al. (2020) worked to deliver the full curriculum, they had to translate and adapt the materials and outcome measures to the home language of participants to aid comprehension. They were therefore given medium WoE B ratings. Schonert-Reichl et al. (2015) omitted three lessons and was given a low WoE B rating.

The cultural adaptation of intervention to the local context is common and can be successful; it can improve participants' understanding, engagement, and the sustainability of intervention effects (Barrera et al., 2017). de Carvalho et al. (2017) and Matsuba et al. (2020) did exactly that, and had documented how they adapted the curriculum to meet the level of understanding of the local participants whose second language is English. Therefore, they were both rated high on WoE C for cultural sensitivity. The remaining three studies did not make any adaptations because it was deemed as unnecessary and were given a medium WoE C rating. Given the heterogeneity of the cultural background in which the interventions took place, there are positive implications for the generalisability of findings.

All studies provided training for teachers on the delivery of MindUP. This improves the external validity of the intervention. In line with the need for cultural adaptation and to work towards good implementation fidelity, de

Carvalho et al. (2017) spent 12.5 hrs longer than the training offered by TGHF, and Matsuba et al. (2020) spent a few extra days so that teachers have enough time to get familiarise with the content and fully understand the curriculum before teaching it.

The length of delivery of MindUP varied as it was dependent on the participants' age in the study. Crooks et al. (2020) delivered weekly lessons that has a duration of 10 to 15 minutes with each core activity lasting at least 1 minute. Thierry et al. (2016) spent 2 to 3 weeks delivering each lesson for 20 to 30 minutes with each core activity lasting from 30 to 60 seconds. Each core activity lasted for 3 minutes in the remaining three studies. Schonert-Reichl et al. (2015) and de Carvalho et al. (2017) delivered a 40 to 60 minutes weekly lesson while Matsuba et al. (2020) delivered the lessons over the course of two school terms. As mentioned previously, de Carvalho et al. (2017) did not assess implementation fidelity while the remaining studies did not describe going beyond assessing surface features. Domitrovich et al. (2008) suggests that the quality of delivery of school-based interventions need to be monitored beyond implementing the framework (e.g. through observations) as it is difficult to assess generalisability through self-report measures like tracking sheets.

Measures

Two studies with younger participants used teacher/parent-rated outcome measures (Crooks et al., 2020; Thierry et al., 2016) while the rest were self-reported outcome measures.

As the current review seeks to evaluate whether MindUP improves internalising problems, it is important that the outcome measures chosen

have construct validity and provides a measure of internalising problems. de Carvalho et al. (2017) used a modified Portuguese version of the PANAS-C (Laurent et al., 1999) which has good convergent and discriminant validity. After analysing, it showed that the negative affect measure was invariant which gave it a high WoE C rating. Schonert-Reichl et al. (2015) used the SPQC (Kusché et al., 1998) which was found to be reliable and valid for the depression subscale with good internal consistency ($\alpha = 0.67$) (Aber et al., 1998). It directly measured internalising problems and was rated high on WoE C.

Matsuba et al. (2020) used the 'anger', 'fear' and 'sadness' outcome measures which makes up the negative affect subdomain of the 'emotion' domain in the NIH Toolbox (Gershon et al., 2013); it was given a medium WoE C rating because it shows test-retest reliability with suitable use in culturally diverse populations and across different developmental stages (Gershon et al., 2013). Crooks et al. (2020) was given a medium WoE C rating for using the internalising problems subscale from the BASC-3 TRS (Reynolds & Kamphaus, 2015). This measure showed internal consistency and discriminant validity (Burback, 2020).

Thierry et al. (2016) was given a low rating on WoE C because the BRIEF-P (Gioia et al., 2002) measured behaviour constructs which was related to executive function in children. Although the emotional control subscale evaluates a child's ability to regulate emotional responses, it does not measure internalising problems and therefore it is not relevant to this review question due to the lack of construct validity.

Findings and effect sizes

Table 7 below provides a description of the outcome measures relevant to assessing internalising problems and its effect sizes. The effect sizes represented by Cohen's d (1992) were all calculated using the Campbell Collaboration Effect Size Calculator (Wilson, n.d.) based on information provided in the studies. As mentioned in the previous section, Thierry et al. (2016) did not use a measure relevant to this review question. Moreover, statistics for the emotional control subscale was unavailable. Therefore, an effect size could not be calculated for this study. Matsuba et al. (2020) reported anger, sadness and fear measures separately, so the overall average change score for negative affect had to be manually calculated by the reviewer before calculating the effect size.

Four studies yielded a small effect size but medium WoE D rating except Schonert-Reichl et al. (2015) that had a high WoE D rating. The average WoE D rating for the five studies is medium at 2.14. It seems that although the studies demonstrated good methodological designs and high quality of evidence, the effectiveness of the intervention on reducing internalising problems is not strong enough to recommend its usage in schools.

A fixed-effects meta-analysis was conducted using the effect sizes from four studies through a software from Meta-Essentials (Suurmond et al., 2017). This retrieved an overall effect size estimate of $d = -0.34$ (95% CI lower = -0.53 , upper = -0.16) which shows that there is a small effect of MindUP on reducing internalising problems with reference to Cohen's (1992) criteria.

Table 8 shows a summary of the details.

Figure 2 shows a forest plot of the four studies. The blue points represent the effect size of each study, while the green point is the overall effect size. The black whiskers are the 95% confidence intervals while the green whiskers represent 95% prediction interval to account for random variation. Although the effect size point estimate was significant overall, the findings showed moderate heterogeneity (Cochran's $Q = 4.42$, $p = 0.220$, $I^2 = 32.07\%$, $\tau^2 = 0.01$) where 32% of variability is due to real differences and not by chance (West et al., 2010, p.21). However, Deeks et al. (2021) suggested that the interpretation of I^2 could be misleading as heterogeneity can be due to several factors and 32% might be negligible and therefore little heterogeneity. Thus, the effect of the intervention on internalising problems appears to differ depending on the context in which the intervention was delivered. The different effect sizes from individual studies seem to support this finding. The culturally adapted MindUP intervention produced smaller effect sizes for de Carvalho et al. (2017) ($d = -0.20$) and Matsuba et al. (2020) ($d = -0.27$) compared to Crooks et al. (2020) ($d = -0.45$) and Schonert-Reichl et al. (2015) ($d = -0.41$). More research needs to be conducted before MindUP can be recommended as a school-based universal intervention for reducing internalising problems.

Table 7

Calculation of effect size for the included studies

Study	Sample size (n)	Measure	Reported outcomes	Effect Size (d) and description	95% Confidence Interval (CI)	WoE D
Crooks et al. (2020)	n = 584	Behavior Assessment System for Children, Third Edition, Teacher Rating Scales (BASC-3 TRS; Reynolds & Kamphaus, 2015) <ul style="list-style-type: none"> Behavioral Symptom Index (BSI) Internalising Problems composite scale	There was a reduction in internalising problems scores in the intervention group compared to the comparison group (adjusted $\beta = -2.73$, $SE = 0.7$, $p < 0.001$).	-0.45 (Small)	-0.62 – 0.29	2.33 (Medium)
de Carvalho et al. (2017)	n = 454	Modified Positive and Negative Affect Schedule for Children (PANAS-C; Laurent et al., 1999). <ul style="list-style-type: none"> Negative affect 	There is a significant decrease in the intervention group compared to the comparison group for negative affect [$F(1, 443) = 4.339$, $p = 0.037$, $d = 0.20$].	-0.20 (Small)	-0.38 – -0.01	1.89 (Medium)
Matsuba et al. (2020)	n = 82	NIH Toolbox for 8 to 12-year olds (Gershon et al., 2013). <ul style="list-style-type: none"> Anger Fear Sadness 	The negative affect score in the intervention group was 0.26(0.72) compared to the comparison group 0.47(0.84).	Average -0.27 (Small)	-0.72 – 0.17	1.72 (Medium)
Schonert-Reichl et al. (2015)	n = 99	Depressive symptoms measured by Seattle Personality Questionnaire for Children (SPQC; Kusché et al., 1988).	There is a significant decrease in the intervention group compared to the control group for expressive suppression [$F(1, 97) = 4.14$, $p = 0.04$, $d = -0.45$].	-0.41 (Small) Value derived from Wilson's (n.d.) effect size calculator	-0.81 – -0.01	2.5 (High)

Thierry et al. (2016)	n = 47	Behavior Rating Inventory of Executive Function-Preschool Version (BRIEF-P; Gioia et al., 2002) <ul style="list-style-type: none"> Emotional control subscale 	No effects were found for the emotional control subscale. No data values were reported.	There is not enough information to calculate effect size.	N.A.	2.28 (Medium)
-----------------------	--------	--	---	---	------	------------------

Note: Cohen's d (1992) was used to show the effect size of the intervention on internalising difficulties. d = 0.2 represents a small effect, d = 0.5 represents a medium effect, and d = 0.8 represents a large effect.

Table 8

Data for the meta-analysis: Sample size (n), effect size (d), variance (V), standard error (SE), 95% confidence interval (CI)

Study	n	d	V	SE	95% CI -	95% CI +
Crooks et al. (2020)	584	-0.45	0.01	0.08	-0.62	0.29
de Carvalho et al. (2017)	454	-0.20	0.01	0.09	-0.38	-0.01
Matsuba et al. (2020)	82	-0.27	0.05	0.22	-0.72	0.17
Schonert-Reichl et al. (2015)	99	-0.41	0.04	0.20	-0.81	-0.01
Overall for four studies	1219	-0.34	-	0.06	-0.53	-0.16

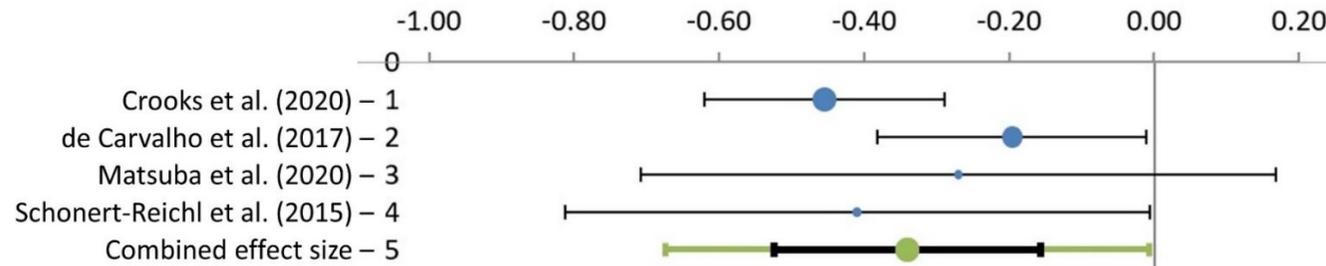


Figure 2. Forest plot of the effect sizes

Conclusions and Recommendations

This review sought to evaluate the effectiveness of MindUP for reducing internalising problems in children ages 3 to 14 years old. Four out of the five studies reviewed were medium quality, and one (Schonert-Reichl et al., 2015) was high quality. Through using the WoE framework (Gough, 2007), Gersten et al.'s (2005) coding protocol and evaluating the statistical effect sizes, the reviewer concludes that there is insufficient evidence to say that MindUP is effective in reducing internalising problems in CYP. In order to conclude that an intervention is evidence based, Gersten et al. (2005) stressed that the overall effect size needs to be significantly more than zero because the effect size value reflects the number of studies that carried out the intervention in question, the number of participants involved, and the uniformity and weight of individual study effects.

In this review, the overall effect size is small; if MindUP was implemented, 34% of the time, a randomly selected individual will benefit from reduced internalising problems. Although this would make a huge difference to that individual child, schools would have to weigh the costs of investment in implementing an intervention that could only impact a small number of their student body. One ongoing study by Kuyken et al. (2017) had set out to address this issue of the effectiveness of MBIs in improving SEMH in CYP and its cost effectiveness.

Studies implementing the culturally adapted MindUP produced smaller effect sizes compared to the ones that implemented it as intended. The task of having to understand and implement MindUP when English is not the teachers' first language could be a disadvantage. It is also possible that the

idea of internalising problems could be understood or expressed differently in different cultures and contexts.

Matsuba et al. (2020) had to train their teachers for longer periods of time as they had no prior knowledge about mindfulness or experience implementing SEL programmes. Additionally, the characteristics of teachers were not reported which could influence how well teachers understood the principles of the intervention and their ability to implement it effectively. Similarly, the theoretical basis of MindUP was new to teachers in the de Carvalho et al. (2017) study. Kabat-Zinn (2003) argues that unless the interventionist's experience with practising mindfulness is grounded in their own personal life, the benefits will not be manifested. The best implementation of Mindful practice is evident when the class teachers are engaged in their own mindful practice or fully understand the process because if the teachers are mindful in the classroom, how they teach would promote focus, resilience and self-soothing skills in their pupils (Gerszberg, 2017).

Limitations

There are three limitations to this review. Firstly, the reviewer had neither considered the variation in age range of participants nor the mode of reporting of (e.g. self-report, teacher-report and parent-report) measures for this review due to the limited number of studies evaluating MindUP to date. These could have contributed to the moderate heterogeneity (West et al., 2010). Secondly, studies included in this review did not explicitly set out to evaluate the impact of MindUP on internalising difficulties. A variety of outcome measures were used to analyse other constructs as well. This limits

the ability to find an effect in relation to internalising problems. Lastly, as the findings from this review only shows a small effect size, there is not enough evidence for EPs to recommend MindUP as an evidence-based universal intervention for internalising difficulties in schools.

Recommendations

MindUP should be evaluated for its effectiveness in improving internalising problems through experimental designs with outcome measures that specifically evaluate internalising difficulties (Liu et al., 2011). Follow-up measures should also be collected to evaluate sustainability of gains beyond post-intervention. Self-report measures to assess the improvement of internalising problems even in CYP as young as 5 years old (Varni et al., 2007) could be used to triangulate evidence from teacher or parent-reports.

Next, as teachers who deliver the mindfulness programmes should themselves be engaged in it (Kabat-Zinn, 2003), so more research could be done to evaluate the impact of conducting the intervention on both teachers' and students' mental wellbeing, and the relationship between the findings.

Additionally, there is a need for future research to continue evaluating the effectiveness of MindUP using the original manualised version in order to find out whether the effectiveness found by Crooks et al. (2020) and Schonert-Reichl et al. (2015) could be strengthened, resulting in a larger effect size.

Lastly, research on the cultural adaptations of MindUP should continue to be evaluated as these interventions have great potential for their efficacy, increased participation, and sustained implementation by local communities (Barrera et al. 2017). It would also be helpful for future

researchers to know that MindUP had recently been rebranded to 'MindUP for life' in February 2021 (Park, 2021).

References

- Aber, J. L., Brown, J. L., Jones, A., & Samples, F. (1998). Seattle Personality Questionnaire 1995.
- Barrera, M., Berkel, C., & Castro, F. G. (2017). Directions for the advancement of culturally adapted preventive interventions: Local adaptations, engagement, and sustainability. *Prevention Science*, 18(6), 640-648. <https://doi.org/10.1007/s11121-016-0705-9>
- Burback, S. (2020). Construct Validity of the Behavior Assessment System for Children-Teacher Rating Scales (BASC-3 TRS): Comparisons with the Adjustment Scales for Children and Adolescents (ASCA).
- CASEL. (2013). *Effective social and emotional learning programs: Preschool and elementary school edition*.
<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwj0leeXuP7vAhW5TBUIHSWGAukQFjACegQIAxAD&url=https%3A%2F%2Fcasel.org%2Fwp-content%2Fuploads%2F2016%2F01%2F2013-casel-guide.pdf&usg=AOvVaw3Bz4SNeMC2-G5sWjS0xRNP>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
<https://doi.org/10.1037/0033-2909.112.1.155>
- Corcoran, R. P., Cheung, A. C., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programmes for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review*, 25, 56-72.
<https://doi.org/10.1016/j.edurev.2017.12.001>

Crooks, C. V., Bax, K., Delaney, A., Kim, H., & Shokoohi, M. (2020). Impact of mindUP among young children: Improvements in behavioral problems, adaptive Skills, and executive functioning. *Mindfulness*, 11(10), 2433-2444. <https://doi.org/10.1007/s12671-020-01460-0>

Davis, M. H. (1983). The effects of dispositional empathy on emotional reactions and helping: A multidimensional approach. *Journal of Personality*, 51(2), 167–184. <https://doi.org/10.1111/j.1467-6494.1983.tb00860.x>

de Carvalho, J. S., Pinto, A. M., & Marôco, J. (2017). Results of a mindfulness-based social-emotional learning program on portuguese elementary students and teachers: a quasi-experimental study. *Mindfulness*, 8(2), 337-350. <https://doi.org/10.1007/s12671-016-0603-z>

Deeks, J.J., Higgins, J.P.T., Altman, D.G. (Eds.). (2021). Chapter 10: Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.2. <https://training.cochrane.org/handbook/current/chapter-10>

Department for Education, & Department of Health. (2015). *Special educational needs and disability code of practice: 0 to 25 years*. <https://www.gov.uk/government/publications/send-code-of-practice-0-to-25>

Department for Education. (2019). *State of the nation 2019: Children and young people's wellbeing research report.*

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/838022/State_of_the_Nation_2019_young_people_children_wellbeing.pdf

Department for Education. (2021). *Statutory framework for the early years foundation stage: Setting the standards for learning, development and care for children from birth to five.*

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/974907/EYFS_framework_-_March_2021.pdf

Department of Health. (2015). *Future in mind: Promoting, protecting and improving our children and young people's mental health and wellbeing.*

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/414024/Childrens_Mental_Health.pdf

Diamond, A. (2009). The interplay of biology and the environment broadly defined. *Developmental Psychology*, 45(1), 1.

<https://doi.org/10.1037/a0014601>

Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., Romanelli, L. H., Leaf, P. J., Greenberg, M. T., & Falongo, N. S. (2008). Maximizing the Implementation Quality of Evidence-Based Preventive Interventions in Schools: A Conceptual

Framework. *Advances in School Mental Health Promotion*, 1(3), 6–28.

<https://doi.org/10.1080/1754730x.2008.9715730>

Dunning, D. L., Griffiths, K., Kuyken, W., Crane, C., Foulkes, L., Parker, J., &

Dalgleish, T. (2019). Research Review: The effects of mindfulness-based interventions on cognition and mental health in children and adolescents—a meta-analysis of randomized controlled trials. *Journal of Child Psychology and Psychiatry*, 60(3), 244-258.

<https://doi.org/10.1111/jcpp.12980>

Durlak, J. A., & Wells, A. M. (1997). Primary prevention mental health programmes for children and adolescents: A meta-analytic review. *American Journal of Community Psychology*, 25(2), 115-152.

<https://doi.org/10.1023/A:1024654026646>

Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH toolbox for assessment of neurological and behavioral function. *Neurology*, 80(11 Suppl. 3), S2–S6.

<https://doi.org/10.1212/WNL.0b013e3182872e5f>

Gersten, R., Fuchs, L. S., Compton, D. L., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71, 149–164.

<https://doi.org/10.1177/001440290507100202>

Gerszberg, C. O. (2017, September 14). The future of education: Mindful classrooms. *Mindful: Healthy mind, healthy life*.

<https://www.mindful.org/future-education-mindful-classrooms/>

Gioia, G. A., Andrews, K., & Isquith, P. K. (1996). Behavior rating inventory of executive function-preschool version (BRIEF-P). Psychological Assessment Resources.

Gioia, G. A., Espy, K. A., & Isquith, P. K. (2002). Behavior Rating Inventory of Executive Function—Preschool Version (BRIEF-P). Odessa, FL: Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). BRIEF: Behavior rating inventory of executive function. Psychological Assessment Resources.

Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education, 22*, 213–228. <https://doi.org/10.1080/02671520701296189>

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. <https://doi.org/10.1037/0003-066X.58.6-7.466>

Gullone, E., & Taffe, J. (2012). The Emotion Regulation Questionnaire for Children and Adolescents (ERQ–CA): A psychometric evaluation. *Psychological Assessment, 24*(2), 409–417. <https://doi.org/10.1037/a0025777>

- Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. *Clinical Psychology: Science and Practice*, 10(2), 144-156. <https://doi.org/10.1093/clipsy.bpg016>
- Kusché, C. A., Greenberg, M. T., & Beilke, R. (1988). *Seattle personality questionnaire for young school-aged children*. [Unpublished manuscript]. University of Washington, Department of Psychology, Seattle.
- Kuyken, W., Nuthall, E., Byford, S., Crane, C., Dalgleish, T., Ford, T., ... & Williams, J. M. G. (2017). The effectiveness and cost-effectiveness of a mindfulness training programme in schools compared with normal school provision (MYRIAD): study protocol for a randomised controlled trial. *Trials*, 18(1), 1-17. <https://doi.org/10.1186/s13063-017-1917-4>
- Kuyken, W., Weare, K., Ukoumunne, O., Vicary, R., Motton, N., Burnett, R., . . . Huppert, F. (2013). Effectiveness of the Mindfulness in Schools Programme: Non-randomised controlled feasibility study. *British Journal of Psychiatry*, 203(2), 126-131.
doi:10.1192/bjp.bp.113.126649
- Laurent, J., Catanzaro, S. J., Joiner, T. E., Rudolph, K. D., Potter, K. I., Lambert, S., & Gathright, T. (1999). A measure of positive and negative affect for children: Scale development and preliminary validation. *Psychological Assessment*, 11(3), 326–338.
<https://doi.org/10.1037/1040-3590.11.3.326>

Lawlor, M. S., Schonert-Reichl, K. A., Gadermann, A. M., & Zumbo, B. D. (2014). A validation study of the mindful attention awareness scale adapted for children. *Mindfulness*, 5(6), 730-741.

<https://doi.org/10.1007/s12671-013-0228-4>

Litvack-Miller, W., McDougall, D., & Romney, D. M. (1997). The structure of empathy during middle childhood and its relationship to prosocial behavior. *Genetic, Social, and General Psychology Monographs*, 123(3), 303–325.

Liu, J., Chen, X., & Lewis, G. (2011). Childhood internalizing behaviour: analysis and implications. *Journal of Psychiatric and Mental Health Nursing*, 18(10), 884-894. <https://doi.org/10.1111/j.1365-2850.2011.01743.x>

Lyubomirsky, S., & Layous, K. (2013). How do simple positive activities increase well-being? *Current Directions in Psychological Science*, 22(1), 57–62. <https://doi.org/10.1177/0963721412469809>

Maloney, J. E., Lawlor, M. S., Schonert-Reichl, K. A., & Whitehead, J. (2016). A mindfulness-based social and emotional learning curriculum for school-aged children: The MindUP Program. In K. A. Schonert-Reichl & R. W. Roeser (Eds.), *Handbook of mindfulness in education: Integrating theory and research into practice* (pp. 313–334). Springer-Verlag Publishing. https://doi.org/10.1007/978-1-4939-3506-2_20

Marsh, H. W., Barnes, J., Cairns, L., & Tidman, M. (1984). Self-Description Questionnaire: Age and sex effects in the structure and level of self-

concept for preadolescent children. *Journal of Educational Psychology*, 76(5), 940–956. <https://doi.org/10.1037/0022-0663.76.5.940>

Matsuba, M. K., Schonert-Reichl, K. A., McElroy, T., & Katahoire, A. (2020). Effectiveness of a SEL/mindfulness program on Northern Ugandan children. *International Journal of School & Educational Psychology*, 1-19. <https://doi.org/10.1080/21683603.2020.1760977>

Maynard, B. R., Solis, M. R., Miller, V. L., & Brendel, K. E. (2017). Mindfulness-based interventions for improving cognition, academic achievement, behavior, and socioemotional functioning of primary and secondary school students. *Campbell Systematic Reviews*, 13(1), 1-144. <https://doi.org/10.4073/CSR.2017.5>

Musty, T. A. (2015). The MindUP curriculum grades 3-5: Brain-focused strategies for learning and living. *School Social Work Journal*, 39(2), 81.

Neff, K. D. (2003). The development and validation of a scale to measure self-compassion. *Self and Identity*, 2(3), 223–250. <https://doi.org/10.1080/15298860309027>

Noam, G. G., & Goldstein, L. S. (1998). *The resilience inventory*. [Unpublished protocol].

Park, J. (2021, February 1). *MindUP rebrands as mindup for life and evolves visual identity to embody new holistic direction*. Little black book.

<https://www.lbbonline.com/news/mindup-rebrands-as-mindup-for-life-and-evolves-visual-identity-to-embody-new-holistic-direction>

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health*, 57, 527–529. <http://dx.doi.org/10.1136/jech.57.7.527>

Reynolds, C. R., & Kamphaus, R. W. (2015). Behavior assessment system for children—third edition (BASC-3). Pearson.

Roeser, R. W., & Zelazo, P. D. (2012). Contemplative science, education and child development: Introduction to the special section. *Child Development Perspectives*, 6(2), 143–145. <https://doi.org/10.1111/j.1750-8606.2012.00242.x>

Schonert-Reichl, K. A., Oberle, E., Lawlor, M. S., Abbott, D., Thomson, K., Oberlander, T. F., & Diamond, A. (2015). Enhancing cognitive and social–emotional development through a simple-to-administer mindfulness-based school program for elementary school children: A randomized controlled trial. *Developmental Psychology*, 51(1), 52–66. <https://doi.org/10.1037/a0038454>

Schonert-Reichl, K. A., Smith, V., Zaidman-Zait, A., & Hertzman, C. (2012). Promoting children's prosocial behaviors in school: Impact of the "Roots of Empathy" program on the social and emotional competence of school-aged children. *School Mental Health: A Multidisciplinary Research and Practice Journal*, 4(1), 1–21. <https://doi.org/10.1007/s12310-011-9064-7>

Sibinga, E. M., Webb, L., Ghazarian, S. R., & Ellen, J. M. (2016). School-based mindfulness instruction: an RCT. *Pediatrics*, 137(1).

<https://doi.org/10.1542/peds.2015-2532>

Sisk, D. A. (2018). The art and science of planting seeds of mindfulness. *Gifted Education International*, 34(2), 118-128.

<https://doi.org/10.1177/0261429417716354>

Song, M. (2003). *Two studies on the Resilience Inventory (RI): Toward the goal of creating a culturally sensitive measure of adolescence resilience*. [Doctoral dissertation]. Harvard Graduate School of Education.

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of Meta-Essentials: a free and simple tool for meta-analysis. *Research synthesis methods*, 8(4), 537-553.

<https://doi.org/10.1002/jrsm.1260>

The Goldie Hawn Foundation. (n.d.). *MindUP*. <https://mindup.org/mindup-for-schools/>

Thierry, K. L., Bryant, H. L., Nobles, S. S., & Norris, K. S. (2016). Two-year impact of a mindfulness-based program on preschoolers' self-regulation and academic performance. *Early Education and Development*, 27(6), 805–821.

<https://doi.org/10.1080/10409289.2016.1141616>

Varni, J. W., Limbers, C. A., & Burwinkle, T. M. (2007). How young can children reliably and validly self-report their health-related quality of life?: An analysis of 8,591 children across age subgroups with the PedsQL™ 4.0 Generic Core Scales. *Health and Quality of Life Outcomes*, 5(1), 1-13. <https://doi.org/10.1186/1477-7525-5-1>

Weare, K. (2019). Mindfulness and contemplative approaches in education. *Current Opinion in Psychology*, 28, 321-326. <https://doi.org/10.1016/j.copsyc.2019.06.001>

Wentzel, K. R. (1993). Motivation and achievement in early adolescence: The role of multiple classroom goals. *The Journal of Early Adolescence*, 13(1), 4-20. <https://doi.org/10.1177/0272431693013001001>

West, S. L., Gartlehner, G., Mansfield, A. J., Poole, C., Tant, E., Lenfestey, N., ... & Lohr, K. N. (2010). Comparative effectiveness review methods: Clinical heterogeneity. <https://www.ncbi.nlm.nih.gov/books/NBK53310/>

Wilson, D. B., Ph.D. (n.d.). Practical Meta-Analysis Effect Size Calculator [Online calculator]. Retrieved February 12, 2021, from <https://campbellcollaboration.org/research-resources/effect-size-calculator.html>

World Health Organization. (n.d.). *Improving the mental and brain health of children and adolescents*. <https://www.who.int/activities/Improving-the-mental-and-brain-health-of-children-and-adolescents>

Zelazo, P. D., & Lyons, K. E. (2012). The potential benefits of mindfulness training in early childhood: A developmental social cognitive neuroscience perspective. *Child Development Perspectives*, 6(2), 154–160. <https://doi.org/10.1111/j.1750-8606.2012.00241.x>

Appendices

Table 1

List of Content in Appendices

Appendix	Content	Page(s)
A	Excluded studies from the review at full text screening and ancestral search with rationale	52-53
B	Weight of Evidence A: Methodological quality	54
C	Changes made to Gersten et al. (2005) coding protocol	55-57
D	Example of a completed coding protocol	58-62
E	Weight of Evidence B: Methodological relevance	63-64
F	Weight of Evidence C: Topic relevance	65-66

Appendix A

Table 2

Excluded studies from the review at full text screening and ancestral search with rationale

Reference of excluded study	Rationale
<p>Arthurson, K. (2015). Teaching mindfulness to year sevens as part of health and personal development. <i>Australian Journal of Teacher Education</i>, 40(5), 2.</p> <ul style="list-style-type: none"> Excluded at full text screening 	<p>Criteria 3: Type of Intervention</p> <ul style="list-style-type: none"> The programme implemented in this study was developed using several different mindfulness programmes including the MindUP curriculum rather than purely implementing MindUP. <p>Criteria 4: Research design and methodology</p> <ul style="list-style-type: none"> This was a pilot study with no comparison or control group.
<p>Johnstone, J. M., Roake, C., Sheikh, I., Mole, A., Nigg, J. T., & Oken, B. (2016). School-based mindfulness intervention for stress reduction in adolescents: Design and methodology of an open-label, parallel group, randomised controlled trial. <i>Contemporary clinical trials communications</i>, 4, 99-104.</p> <ul style="list-style-type: none"> Excluded at full text screening 	<p>Criteria 3: Study did not deliver MindUP</p> <ul style="list-style-type: none"> The programme implemented in this study was developed using several different mindfulness programmes. Only the MindUP lessons involving brain science were delivered rather than the full programme.
<p>Kielty, M., Gilligan, T., Staton, R., & Curtis, N. (2017). Cultivating mindfulness with third grade students via classroom-based interventions. <i>Contemporary School Psychology</i>, 21(4), 317-322.</p> <ul style="list-style-type: none"> Excluded at full text screening 	<p>Criteria 3: Study did not deliver MindUP</p> <ul style="list-style-type: none"> The programme implemented in this study was developed using several different mindfulness programmes including the MindUP curriculum rather than purely implementing MindUP. <p>Criteria 4: Research design and methodology</p> <ul style="list-style-type: none"> A mixed methods design was used with no comparison or control group.
<p>Schonert-Reichl, K. A., & Lawlor, M. S. (2010). The effects of a mindfulness-based education program on pre-and early</p>	<p>Criteria 3: Study did not deliver MindUP</p> <ul style="list-style-type: none"> Although CASEL (n.d.) and Maloney et al. (2016) described that this paper delivered

<p>adolescents' well-being and social and emotional competence. <i>Mindfulness</i>, 1(3), 137-151.</p> <ul style="list-style-type: none">• Excluded at ancestral search full text screening	<p>'MindUP', the intervention described within the paper was a different one called Mindfulness Education (ME) program, and "teachers underwent an intensive 1-day training session and received bi-weekly consultation from one of the authors of the ME program curriculum (Nancy Fischer)." (Schonert-Reichl & Lawlor, 2010, p. 143).</p>
<p>Schonert-Reichl, K. A., Oberle, E., Lawlor, M. S., Abbott, D., Thomson, K., Oberlander, T. F., Diamond, A. (2010). Accelerating cognitive and social emotional development in elementary school classrooms: Benefits of a simple to administer program. Manuscript submitted for publication (copy on file with author).</p> <ul style="list-style-type: none">• Excluded at ancestral search full text screening	<p>Having contacted the first author, this paper was published with the reference below and is already included for this review.</p> <p>Schonert-Reichl, K. A., Oberle, E., Lawlor, M. S., Abbott, D., Thomson, K., Oberlander, T. F., & Diamond, A. (2015). Enhancing cognitive and social-emotional development through a simple-to-administer mindfulness-based school program for elementary school children: A randomised controlled trial. <i>Developmental psychology</i>, 51(1), 52. https://doi.org/10.1037/a0038454</p>

Appendix B

Weight of Evidence A: Methodological quality

Table 9 below shows the criteria to be met for each respective WoE A rating. It has been adapted from Gersten et al. (2005). A study that is given a WoE A rating of 'high' is also considered 'high quality' by Gersten et al. (2005). A study that is given a 'medium' rating is considered as 'acceptable quality' by Gersten et al. (2005). A study is given a 'low' rating by the reviewer as it did not meet the 'high' nor 'acceptable' qualities suggested by Gersten et al. (2005).

Table 3

WoE A Rating Criteria Adapted from Gersten et al. (2005)

WoE A rating	Criteria
3 High	9 or more Essential Quality Indicators met and 5 or more Desirable Quality Indicators met
2 Medium	9 Essential Quality Indicators met and 1 to 4 Desirable Quality Indicators met
1 Low	Less than 9 Essential Quality Indicators met (regardless of number of Desirable Quality Indicators met)

Appendix C

Changes made to Gersten et al. (2005) coding protocol

Changes made to the coding protocol (Gersten et al., 2005) are annotated in the following manner: Added texts are [written within square brackets].

Deleted texts have ~~strikes across them~~. The rationales have been written in *italics and underlined*. The rest of the text is presented in its original form.

Due to the changes, there are now 10 desirable criteria instead of 8.

According to Gersten et al. (2005), in order to be rated as 'high' quality, a study must have met 50% of the desirable criteria. Therefore, a study must now meet 5 desirable criteria instead of 4, also reflected in Appendix D, Table 4.

Table 4

Essential and Desirable Quality Indicators for Group Experimental and Quasi-Experimental Research Articles and Reports

Essential Quality Indicators

Quality Indicators for Describing Participants

- | | |
|---|--|
| 1 | Was sufficient information provided to [indicate and explain why certain participants were excluded from the intervention]? determine/confirm whether the participants the disability(ies) or difficulties presented? |
|---|--|

Rationale: As this systematic literature review is looking at whether MindUP, as a universal intervention, can prevent or reduce Social, Emotional and Mental Health (SEMH) difficulties, it would be more important if the study indicated whether any participants were excluded from the intervention and the reasons why they were excluded.

- | | |
|---|--|
| 2 | Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions? |
|---|--|
-

- | | |
|---|--|
| 3 | Was sufficient information given characterising the interventionists or teachers provided? Did it indicate whether they were comparable across conditions? |
|---|--|
-

Quality Indicators for Implementation of the Intervention and Description of Comparison Conditions

EFFECTIVENESS OF MINDUP

1	Was the intervention clearly described and specified?
2	Was the fidelity of implementation described and assessed?
3	Was the nature of services provided in comparison conditions described?

Quality Indicators for Outcome Measures

1	Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance?
2	Were outcomes for capturing the intervention's effect measured at the appropriate times?

Quality Indicators for Data Analysis

1	Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study?
2	Did the research report include not only inferential statistics but also effect size calculations?

Desirable Quality Indicators

1	Was data available on attrition rates among intervention samples? Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%? <u><i>Rationale: Questions were separated as they were judged to be addressing different issues.</i></u>
2	Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?
3	Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures? Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions? <u><i>Rationale: Questions were separated as they were judged to be addressing different issues.</i></u>
4	Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?
5	Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?
6	Was evidence of the criterion-related validity and construct validity of the measures provided?
7	Did the research team assess not only surface features of fidelity implementation (e.g., number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?

EFFECTIVENESS OF MINDUP

-
- 8 Was any documentation of the nature of instruction or series provided in comparison conditions?
-
- 9 Did the research report include ~~actual audio or videotape excerpts~~ [examples of paper work such as lesson plans or worksheets] that capture the nature of the intervention?
- Rationale: Some studies included appendices with lessons plans which captures the full nature of the intervention. This was judged to provide similar ecological validity to audio or videotape excerpts.*
-
- 10 Were results presented in a clear, coherent fashion?
-

Appendix D

Example of a completed coding protocol

Coding protocol: Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C, & Innocenti, M. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71,149-164.

Reference for this study: Crooks, C. V., Bax, K., Delaney, A., Kim, H., & Shokoohi, M. (2020). Impact of mindUP among young children: Improvements in behavioral problems, adaptive Skills, and executive functioning. *Mindfulness*, 11(10), 2433-2444.

Table 5

Essential and Desirable Quality Indicators for Group Experimental and Quasi-Experimental Research Articles and Reports

Essential Quality Indicators

Quality Indicators for Describing Participants

- 1 Was sufficient information provided to indicate and explain why certain participants were excluded from the intervention?

Yes – Figure 1 in the paper shows how the final group of participants were chosen.

No

Unknown/Unable to Code

- 2 Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?

Yes

No – Non-randomised allocation was used.

Unknown/Unable to Code

- 3 Was sufficient information given characterising the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?

Yes

No

Unknown/Unable to Code

EFFECTIVENESS OF MINDUP

Quality Indicators for Implementation of the Intervention and Description of Comparison Conditions

1 Was the intervention clearly described and specified?

- Yes
- No
- Unknown/Unable to Code

2 Was the fidelity of implementation described and assessed?

- Yes
- No
- Unknown/Unable to Code

3 Was the nature of services provided in comparison conditions described?

- Yes
- No
- Unknown/Unable to Code

Quality Indicators for Outcome Measures

1 Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance?

- Yes
- No
- Unknown/Unable to Code

2 Were outcomes for capturing the intervention's effect measured at the appropriate times?

- Yes
- No
- Unknown/Unable to Code

Quality Indicators for Data Analysis

1 Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study?

EFFECTIVENESS OF MINDUP

- Yes – Teacher reported measures were provided.
 - No
 - Unknown/Unable to Code
-

2 Did the research report include not only inferential statistics but also effect size calculations?

- Yes – Intra-class correlations (ICC) were reported.
 - No
 - Unknown/Unable to Code
-

Desirable Quality Indicators

1 Was data available on attrition rates among intervention samples?

- Yes – The participants and interventionists were present throughout the intervention so there is no attrition to report on.
 - No
 - Unknown/Unable to Code
-

2 Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?

- Yes
 - No
 - Unknown/Unable to Code
-

3 Did the study provide not only internal consistency reliability but also test–retest reliability and interrater reliability (when appropriate) for outcome measures?

- Yes
 - No
 - Unknown/Unable to Code
-

4 Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?

- Yes
 - No – Teachers who implemented the intervention were engaged as paid research assistants and rated their pupils in their classes.
 - Unknown/Unable to Code
-

EFFECTIVENESS OF MINDUP

-
- 5 Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?
- Yes
 No
 Unknown/Unable to Code
-
- 6 Was evidence of the criterion-related validity and construct validity of the measures provided?
- Yes
 No
 Unknown/Unable to Code
-
- 7 Did the research team assess not only surface features of fidelity implementation (e.g., number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?
- Yes
 No
 Unknown/Unable to Code
-
- 8 Was any documentation of the nature of instruction or series provided in comparison conditions?
- Yes – Topics from the regular curriculum were mentioned.
 No
 Unknown/Unable to Code
-
- 9 Did the research report include examples of paper work such as lesson plans or worksheets that capture the nature of the intervention?
- Yes
 No
 Unknown/Unable to Code
-
- 10 Were results presented in a clear, coherent fashion?
- Yes
 No
 Unknown/Unable to Code
-

EFFECTIVENESS OF MINDUP

Calculation of WoE A Rating:

	(3 = High)	(2 = Medium)	(1 = Low)
	9 or more Essential Quality Indicators met and 5 or more Desirable Quality Indicators met	9 Essential Quality Indicators met and 1 to 4 Desirable Quality Indicators met	Less than 9 Essential Quality Indicators met (regardless of number of Desirable Quality Indicators met)
Essential Quality Indicators met (maximum = 10)	9		
Desirable Quality Indicators met (maximum = 10)		5	
Overall rating (1 to 3) = 3			

Appendix E

Weight of Evidence B: Methodological relevance

WoE B determines the quality, appropriateness, and methodological relevance of each study in answering the review question (Gough, 2007). As suggested by findings from Petticrew and Roberts (2003) about 'effectiveness' questions, the criteria laid out in Table 10 below is suited to make judgements on the 'effectiveness' of MindUP.

Table 10

WoE B Rating Criteria

	(3 = High)	(2 = Medium)	(1 = Low)	Rationale
Selection of participants	Randomised Controlled Trial with random assignment of participants to intervention and control groups	Quasi-experimental design with an element of random assignment to ensure groups were equal	Quasi-experimental design without any random assignment	Selecting the participants randomly prevents selection bias
Intervention delivered	Full intervention curriculum was delivered as intended	Intervention curriculum was adapted	Intervention was delivered with 20% or less omitted	To fully evaluate the effectiveness of MindUP
Comparison group intervention	Clearly described the comparison group condition and point out how it differs from the intervention condition	Clearly described the comparison group condition but did not point out how it differs from the intervention condition	Did not provide any information about the comparison group condition	To find out whether improvement in the intervention group was due to MindUP

EFFECTIVENESS OF MINDUP

Table 11

*Summary of WoE B
Rating*

	Selection of participants	Intervention delivered	Comparison group intervention	WoE B rating
Crooks et al. (2020)	1	3	2	2
de Carvalho et al. (2017)	2	2	1	1.67
Matsuba et al. (2020)	2	2	1	1.67
Schonert-Reichl et al. (2015)	3	1	2	2
Thierry et al. (2016)	1	3	3	2.33

Note. 1.4 or less = Low, 1.5 to 2.4 = Medium, 2.5 to 3 = High.

Appendix F

Weight of Evidence C: Topic relevance

Table 12

WoE C Rating Criteria

	(3 = High)	(2 = Medium)	(1 = Low)	Rationale
The intervention seeks to address internalising problems in children	An outcome measure that directly measures internalising problems	An outcome measure with subscale measuring internalising problems	Outcome measures do not particularly measure internalising problems	To fully evaluate the effectiveness of MindUP for reducing internalising problems in children
Cultural sensitivity	Intervention adapted to the culture to which it is delivered including translation of materials to the local language	Intervention delivered without any adaptations as it was deemed as not required	Intervention delivered without any adaptations although it was needed	To ensure that children fully understand the intervention that was delivered which facilitates their full participation

EFFECTIVENESS OF MINDUP

Table 13

Summary of WoE C Rating

	The intervention seeks to address internalising problems in children	Cultural sensitivity	WoE C rating
Crooks et al. (2020)	2	2	2
de Carvalho et al. (2017)	3	3	3
Matsuba et al. (2020)	2	3	2.5
Schonert-Reichl et al. (2015)	3	2	2.5
Thierry et al. (2016)	1	2	1.5

Note. 1.4 or less = Low, 1.5 to 2.4 = Medium, 2.5 to 3 = High.