Doctorate in Educational and Child Psychology *Ghowrigah Parameswaran*

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

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

How effective are mandala interventions at reducing test anxiety in school age children?

Summary

Evidence suggests that students are experiencing increasing levels of anxiety which impede their ability to learn and flourish in education (Merikangas et al., 2010). In particular, test anxiety is becoming an increasing concern for many pupils (Ergene, 2003). Test anxiety is characterised by negative thoughts and feelings towards test-taking situations. Pupils experiencing test anxiety may feel emotions such as tension, stress and worry. With the expectations of schools to provide social, emotional and mental health (SEMH) support for children and young people (CYP; DfE, 2014), this review explored the effectiveness of mandala interventions at reducing test anxiety. In this review, 'mandala interventions' refer to brief mindfulness-based activities involving pupils drawing or colouring mandalas. A total of five studies were identified and critically appraised in this review. Effect sizes were calculated, showing small and medium size effects. Given the simplicity and feasibility, mandala interventions hold promise as a viable method for reducing test anxiety in children. Future research may wish to address the limitations discussed in this review. For instance, participants in the included studies were aged between nine and fourteen years, therefore, further research is needed to explore whether similar effects are found in older children who experience more regular testing.

Introduction

Test Anxiety

It is believed that school-related stressors such as increased academic pressures and high-stakes testing contribute to the increasing prevalence of anxiety in CYP (Pope, 2010). One form of anxiety that is predicted to rise is test anxiety; this can be defined as a mixture of state anxiety, the perceived threat before a test, and the worry of performance-evaluative situations (Lowe et al., 2008). This differs from general anxiety, which can be referred to as an unpleasant emotional state without a specific object (Reber, 1995, as cited in Putwain, 2008). Test anxiety is a multifaceted construct consisting of cognitive, emotional, behavioural and physiological components (Sarason, 1984). In the literature, test anxiety has been conceptualised in three different ways: as a personality trait, an emotional state and comparable to a clinical disorder (Putwain, 2008). Additionally, there has been debate around the subject-specificity of test anxiety (Sapp et al., 1995), which indicates a lack of consensus regarding test anxiety as a construct.

Despite the multidimensional nature of test anxiety, early test inventories did not include distinct scales to measure the different components (Putwain, 2008). In fact, traditional test anxiety measures emphasise a two-factor model which mainly assess cognitive and emotionality symptoms (von der Embse et al., 2013). Furthermore, in recent research, the focus of test anxiety has been on the cognitive and physiological elements of the construct (McDonald, 2001). The cognitive component relates to an individual's concern about their performance, whilst the physiological component or emotionality refers to negative affective and physiological symptoms.

Putwain and Daly (2014) found that in their sample of 2,345 secondary school students, 16% of participants reported themselves to be highly test anxious. This is concerning as high levels of test anxiety can negatively affect student performance on state-wide achievement tests (Segool et al., 2013). It is likely that a proportion of students are underperforming in academic tests due to test anxiety, which is particularly problematic as performance in high-stakes examinations can determine important outcomes for students (von der Embse et al., 2013).

Mandala Interventions

There are several psychologically-informed interventions to reduce test anxiety. Examples include Cognitive Bias Modification (Sportel et al., 2013), and TestEdge, a self-regulation programme (Bradley et al., 2010). However, despite being underpinned by psychological theories, these interventions can often take weeks to implement and therefore are less likely to be fully accepted by schools (Fridrici & Lohaus, 2009). Therefore, there is a need for a feasible intervention that school staff can straightforwardly implement to reduce test anxiety.

Mandala interventions appear to be a promising method for reducing test anxiety. Originating from Eastern culture, a mandala is a circular design which is believed to promote psychological healing when created by an individual (Henderson et al., 2007). Mandala interventions can be considered mindfulness-based as individuals are found to remain focused and aware of the present moment during the activity (Barrett, 2015). Structured mandalas, which are composed of symmetrical shapes, are believed to promote focused attention and awareness of the present moment (Beckwith, 2014) thus promoting a mindful state which is thought to reduce anxiety (Curry & Kasser, 2005).

Mandala interventions can reduce test anxiety by targeting both the cognitive and physiological components. The cognitive component can be described as worry over how test performance will be perceived by others (Putwain, 2008). It is often associated with pupils experiencing self-deprecating thoughts about their abilities, which can lead to physiological symptoms of anxiety (Bradley et al., 2010). Considering the mindful nature of mandala interventions and the focused attention it can bring, this intervention can help pupils to focus on the present moment and make adaptive changes in their thought patterns (Teasdale et al., 1995). This can lead to individuals paying less attention to the intrusive thoughts which can worsen feelings of test anxiety and jeopardize performance (Sarason, 1984). Additionally, it is presumed that by changing thoughts related to the testing situation, pupils develop better control of their anxiety-related emotions, which consequently improves test performance (Spielberger & Vagg, as cited in Bradley et al. 2010).

Mandala interventions can also target the physiological component of test anxiety, which can manifest itself in symptoms such as muscle tension and increased heart rate (McDonald, 2001). It is generally accepted that brain regions such as the amygdala and prefrontal cortex are associated with anxiety (Edenfield & Saeed, 2012). For instance, the amygdala is the area of the brain which triggers the fight or flight response in order to protect the body from danger (Breedlove & Watson, 2013). This response can be activated by everyday stressors such as preparing for a test or being late for an appointment (Napoli et al., 2005). Research is beginning to show that engaging in mindfulness practice can lead to changes in specific brain regions. This includes the deactivation of the amygdala (Goldin & Gross, 2010), a decrease in grey matter volume in the right amygdala (Taren et al., 2013), greater cortical thickness in the hippocampus (Hernández et al., 2016) which is associated with

improved test performance (Mackey et al., 2015), and the activation of the parasympathetic nervous system (PNS) which decreases physiological arousal (Morrell, 2018). This indicates that the mindful nature of mandala interventions can circumvent the physiological stress response inherent in test anxiety by activating the PNS and thus inducing a relaxation state which is more conducive to improved test performance. Furthermore, research involving mandala colouring more specifically indicate the reduction of physiological symptoms of anxiety post intervention (Rose & Lomas, 2020).

It should be noted however that many of the mindfulness interventions in these research studies are long-term programmes. Further research is needed in exploring the effectiveness of mandala interventions at targeting the cognitive and emotionality components specifically as there is a paucity of research in this area. Additionally, more attention needs to be given to the test anxiety measures so that they include reliable measures of the cognitive and affective components. This will allow firmer conclusions to be drawn regarding the effectiveness of mandala interventions.

Psychological Relevance

Although mindfulness is rooted in Eastern traditions (Bluth & Blanton, 2014), mindfulness-based interventions such as mandala colouring appear to have clear parallels with positive psychology, which is primarily focused on psychological well-being (Marques et al., 2011). In particular, both mindfulness and art creation can be linked to the psychological theory of Flow (Csikszentmihalyi, 1975). Flow is a psychological state that occurs when individuals become fully immersed in an enjoyable activity (Csikszentmihalyi, 2002). A main condition for experiencing flow is the perceived challenges of the task or activity, and the skills possessed by the

individual to manage the challenges. Similar to mindfulness, a state of flow requires concentration and focused attention, both of which are thought to be present during mandala interventions (Barrett, 2015). As mandala interventions can facilitate in-depth attention and engagement, it is not unexpected that short-term mandala interventions have been found to significantly improve the flow state (Chen et al., 2019). Although the flow experience is often associated with positive mood, research has found that the experience of flow can negatively predict anxiety (Mao et al., 2020). It is therefore possible that the intrinsic focus and positive feelings inherent in a state of flow are effective in reducing feelings of anxiety. Furthermore, the focused attention characteristic of a flow state may be especially beneficial for reducing test anxiety in pupils as they are likely to experience difficulties with attention (Swanson & Howell, 1996).

Although it could be argued that mindfulness-based interventions are lacking in explicit psychological theory, they may be considered as practice-based evidence if they have high feasibility and are recognised as effective interventions (Barkham, 2010 as cited in Soares & Woods, 2020). Furthermore, as the psychological profession continues to expand and evolve, emerging interventions such as those based in mindfulness may become more explicitly integrated into psychological theory and practice in the future (Soares & Woods, 2020).

Relevance to school settings

Statutory guidance has made it clear that schools are required to make reasonable adjustments for CYP with special educational needs and disabilities (SEND). Additionally, the importance of multidisciplinary working is emphasised in order to meet their individual needs (SEND Code of Practice, 2015).

However, given the increasing budget cuts to Child, Adolescent Mental Health Services (Thorley, 2016), there is evidently increasing pressure on schools to provide mental health support for CYP. Furthermore, as the prevalence of student test anxiety is expected to increase (Wren & Benson, 2004), there is a need for educational staff and Educational Psychologists (EPs) to judge the risk posed by test anxiety for individual students and plan intervention to ameliorate against its negative impact (Putwain & Daly, 2014). As EPs hold knowledge of evidence-based interventions and practice-based evidence, they are well placed to support schools to implement suitable interventions to support the wellbeing of CYP.

Review Question

How effective are brief mandala interventions at reducing test anxiety in school age children?

Critical Review of the Evidence

Literature Search

A systematic literature search was conducted on 22nd January 2021. The search terms and databases used can be found in Table 1. Due to the exceptionally large number of results on PsycINFO (n > 18,000), a filter was applied to include only those that had the search terms in the abstract. The search yielded a total of 63 results, of which 16 were duplicates. All duplicates and a further 42 results were removed after the title and abstracts were screened according to the inclusion and exclusion criteria (see Table 2). A total of 6 journal articles and theses were read in full, and one was excluded with rationale. A list of the studies excluded at the abstract and full article reading stage can be found in Appendix A. The flow chart below (Figure 1) outlines the systematic searching process. A total of 5 studies were included in this review (see Table 3). These are described in further detail in Appendix B.

Table 1

List of search terms used in the database search

Databases	Search terms
Web of Science Core Collection	(children *OR* teenagers *OR* adolescents *OR* "school age" *OR* "primary age" *OR* pupils)
	AND
ERIC	("test anxiety" *OR* "examination anxiety" *OR* "exam anxiety")
PsycINFO	*AND*
	(mandala *OR* "mindful colouring" *OR* "mindful coloring" *OR* drawing *OR* art *OR* colouring *OR* coloring)

Table 2

Criteria for Inclusion with Rationale

		Inclusion criteria	Exclusion criteria	Rationale
1	**Type of publication**	Journal articles and unpublished theses or dissertations	Publications that are not journal articles or dissertations and theses e.g. newspapers, books and reports	To ensure the search for empirical data was as complete as possible in order to reduce the risk of publication bias
2	**Publication date**	Studies published from 2011 onwards	Studies published before 2011	This review aims to evaluate the research from within the last ten years.
3	**Design**	Empirical research that used experimental design with pre and post measures	Research with no pre and post measures, such as qualitative research, and meta-analyses or systematic literature reviews.	To evaluate test anxiety levels pre and post intervention.
4	**Context and Language**	Studies from any geographical contexts that are published in English.	Studies that are not produced in English	The researcher does not have access to resources to translate studies into English.
5	**Intervention**	Intervention is art-based and involves pupils colouring or drawing their own mandalas.	Interventions that do not involve participants colouring or drawing their own mandalas.	This review is looking at the effectiveness of colouring or drawing mandalas on test anxiety.
6	**Participants**	Participants are of English compulsory school age (5-16)	Participants are younger than 5 or older than 16 years of age.	This review aims to evaluate the effectiveness of interventions that can be applied in school settings.
7	**Outcome variable**	Study has an outcome variable measuring test anxiety	Study has outcome variables which do not measure test anxiety	This review is looking at the reduction of test anxiety.

Figure 1

Flow Chart of Search Results

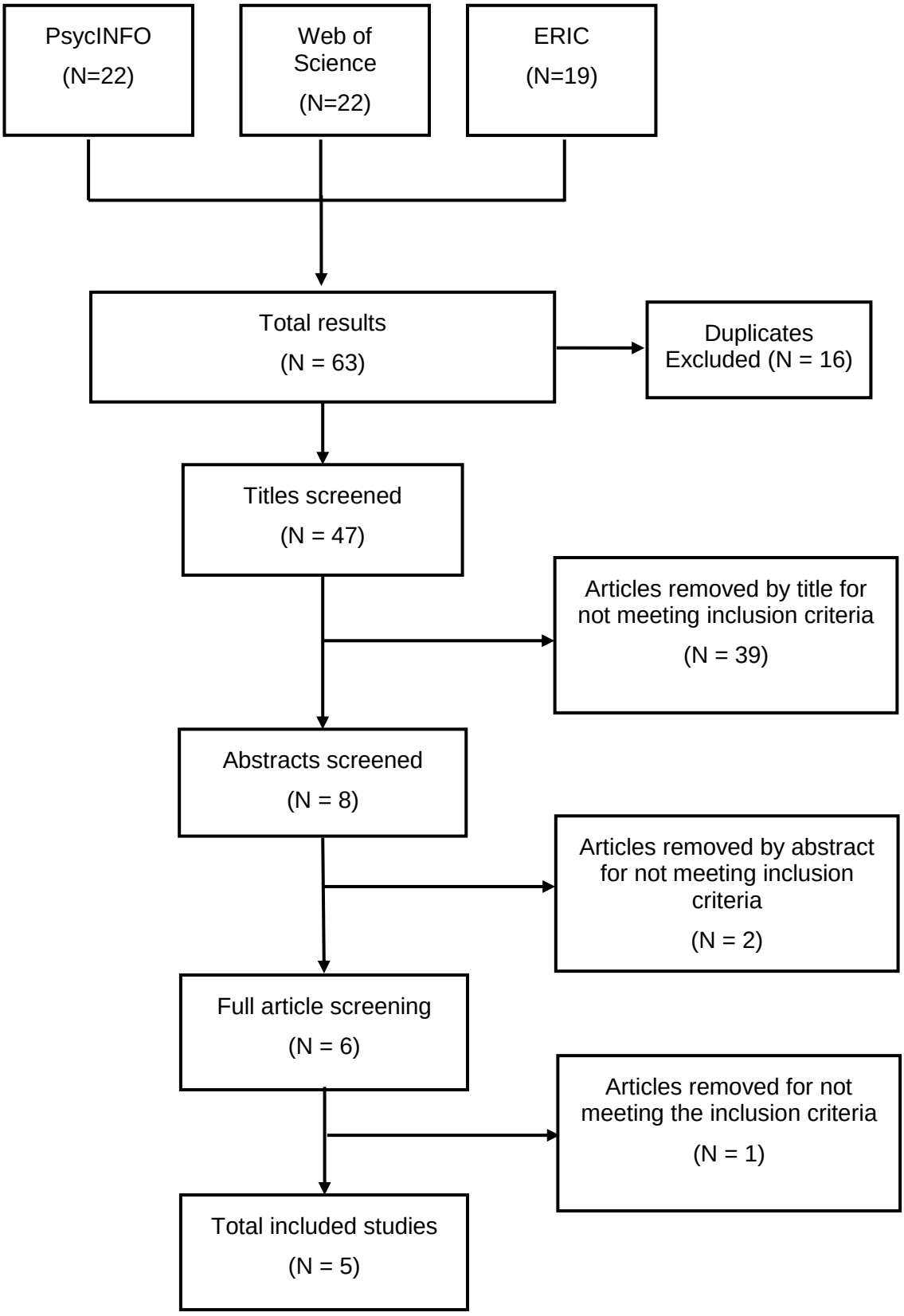

Table 3

List of studies included in the review

Included studies
1. Campbell, A. C. (2012). *Reduction of test anxiety by using mandalas: A pilot study* (Order No. 1519269). Available from ProQuest Dissertations & Theses Global
2. Carsley, D., Heath, N. L., & Fajnerova, S. (2015). Effectiveness of a classroom mindfulness coloring activity for test anxiety in children. *Journal of Applied School Psychology, 31*(3), 239-255.
3. Morrell, M. E. (2018). *Test Anxiety Reduction Through Brief Mindfulness Meditation and Mandala Coloring: School-Based Mindfulness Interventions for Urban Youth* (Order No. 10934041). Available from ProQuest Central; ProQuest Dissertations & Theses Global.
4. Carsley, D., & Heath, N. L. (2018). Effectiveness of mindfulness-based colouring for test anxiety in adolescents. *School Psychology International, 39*(3), 251-272.
5. Carsley, D., & Heath, N. L. (2019). Evaluating the effectiveness of a mindfulness coloring activity for test anxiety in children. *The Journal of Educational Research, 112*(2), 143-151.

Quality and Relevance of Selected Studies

Gough's (2007) Weight of Evidence (WoE) framework was used to critically appraise the five studies. This framework has three main dimensions; WoE A, WoE B and WoE C. The mean value of these dimensions is calculated to establish the overall assessment of each study, providing a WoE D score. WoE A judgements relate to the methodological quality of the studies, WoE B judgements relate to the appropriateness of the studies' methodology in answering this particular review question, and WoE C judgements relate to the appropriateness of the studies in answering the review question. Table 4 below shows the WoE data for all 5 studies. The Gersten et al. (2005) protocol was

adapted (see Appendix D) and used to critically appraise the methodological quality of the studies. Completed coding protocols for each study can be found in Appendix E.

Table 4

Weight of Evidence scores

Research	WoE A	WoE B	WoE C	WoE D
Campbell (2012)	1 Low	1 Low	1.7 Low	1.2 Low
Carsley et al. (2015)	2 Medium	3 High	2.5 Medium	2.5 Medium
Morrell (2018)	1 Low	3 High	2.3 Medium	2.1 Medium
Carsley & Heath (2018)	2 Medium	3 High	2.2 Medium	2.4 Medium
Carsley & Heath (2019)	2 Medium	3 High	2.3 Medium	2.4 Medium

Please note: A score of 1 -1.9 is low, 2 – 2.9 is medium and a score of 3 is high.

Participants

A total of 446 participants, aged nine to fourteen years, took part in the included studies. Three studies reported the mean age of participants and the standard deviations (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019), resulting in a high rating for the WoE C criterion, 'Age of Participants'. Morrell (2018) reported the mean age of participants which resulted in a medium rating for the same criterion. Although Campbell (2012) did not provide a mean age, the age range was reported to be between nine and twelve years. This therefore increases the generalisability of the findings from this review to children of school age in the United Kingdom (UK), as all participants were also of school age.

Three studies took place in Montreal, Canada (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019), one study took place in Illinois,

Chicago (Morrell, 2012), and one study took place in Florida, USA (Campbell, 2012). All studies were deemed to be relatively applicable to the UK education system, resulting in a medium rating for the WoE C criterion 'Location'.

All studies used convenience sampling to recruit participants. For example, in four studies, the researchers distributed consent forms for students to take home to their parents (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019), or disseminated them directly to parents during drop off times (Morrell, 2018). These studies did not report selecting participants with anxiety disorders or difficulties, rather, participants were selected based on their ages, which varied by study. Participants in the Campbell (2012) study were referred by their maths teacher based on displays of test anxious behaviours. It was not reported whether any of the participants in the study had a diagnosed anxiety disorder, although it indicated that participants were not taking any anxiety medication.

Information about the ethnic backgrounds of the participants was only collected in the Morrell (2018) study. This sample in this study comprised of Hispanic students only, which resulted in a medium WoE C rating for criterion 'Ethnicity of Participants'. No studies received a high rating for this criterion as the samples were not comprised of participants from at least three different ethnic backgrounds. Campbell (2012) did not report the ethnic backgrounds of participants, and participants in the remaining studies (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019) were predominantly Caucasian.

Design

Four studies (Carsley et al., 2015; Morrell, 2018; Carsley & Heath, 2018; Carsley & Heath, 2019) used a randomised control trial (RCT) design. This is considered the gold standard design for evaluating treatments (Cochrane Collaboration, 2011). Additionally, the randomisation utilised in RCTs can reduce bias and provide a rigorous tool to explore causal relationships between the independent and dependent variable (Hariton & Locascio, 2018). Due to this, these studies received a 'high' WoE B rating. The studies with an RCT design had active control groups where participants engaged in free drawing or colouring (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019), or independently read short stories (Morrell, 2018). Morrell (2018) included a second intervention group where participants listened to a guided mindfulness meditation script, however, for the purpose of this review, only information regarding the mandala intervention group and the control group was considered.

Campbell (2012) used a quasi-experimental pretest-posttest design with six participants and no control group. The participants selected in these designs often represent a purposeful sample, therefore, results from the findings can only be confidently applied to the selected participants rather than the general population (Stratton, 2019). A further critique of this design are the threats to internal validity, such as reactivity of measurement or endogenous change (Barker, Pistrang & Elliot, 2015). Findings must therefore be interpreted cautiously as external factors could have affected the outcome measure. Due to this, this study received a 'low' WoE B rating.

All studies used various methods to induce test anxiety, contributing to a higher WoE A score. Three studies introduced a test within the method (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019), and one study used deception (Morrell, 2018) to induce test anxiety. In contrast, the Campbell (2012) study took place two weeks prior to the Florida Comprehensive Assessment Test (FCAT), and participants were not subject to researcher-led methods to induce anxiety.

Intervention

Given the brief nature of the mandala intervention, four of the studies took no longer than an hour in total (Carsley et al., 2015; Morrell, 2018; Carsley & Heath, 2018; Carsley & Heath, 2019), and one study was conducted over the course of four weeks (Campbell, 2012). In all studies participants engaged in a mandala intervention for 10 – 15 minutes, however, the nature of the interventions varied slightly. Participants in Campbell (2012) were briefly taught about mandalas, which were referred to as any art form executed within a circular shape (Henderson et al., 2007). They created their own mandalas by drawing a circle and filling it in with as much detail as they liked. It could be said that this task lacked the structured component required for the task to be mindful (Greenberg & Harris, 2012). As the intervention activity in Campbell (2012) was described clearly, and examples of paperwork was provided to allow for replicability, a 'high' WoE C score was given for 'Intervention Description'. In the remaining studies (Carsley et al., 2015; Morrell, 2018; Carsley & Heath, 2018; Carsley & Heath, 2019), participants coloured structured mandalas for 10 – 15 minutes. Morrell (2018) and Carsley et al. (2016) provided clear instructions and an example of the

structured mandala that was supplied, resulting in a high WoE C score for 'Intervention Description'.

In all studies, the intervention was delivered by the researchers in schools, resulting in a medium WoE C score for 'Intervention Delivery'. The studies did not receive a high score as the interventions were not delivered by school staff; having school staff deliver the intervention could have further increased the external validity of the findings.

Measures

All studies measured test anxiety at pre and post using self-report measures. Campbell (2012) used the 30-item Children's Test Anxiety Scale (CTAS; Wren & Benson, 2004). The reliability of the CTAS was reported in the paper to be 0.92, however, test-retest reliability was not provided, resulting in WoE A penalties. Campbell (2012) also gathered qualitative data in the form of meeting observation notes. However, given that this review is exploring the effectiveness of colouring mandalas, only data from the CTAS was considered. The remaining studies used the State-Trait Anxiety Inventory for Children State form (STAIC-S; Spielberger et al., 1973).

All studies varied with how much information they reported regarding the reliability of the measures. For example, two studies (Carsley & Heath, 2018; Carsley & Heath, 2019) used multiple measures, and although they reported the internal consistency reliability and test-retest reliability of the STAIC and their mindfulness attention measure, they did not include information about the test-retest reliability of their third measure. Additionally, none of the

studies reported information relating to the criterion-related validity or construct validity, resulting in further WoE A penalties.

Participants in four of the studies (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019; Morrell, 2018) completed the post-intervention measure immediately after the intervention, resulting in a high WoE C rating for 'Outcome Measures'. The Campbell (2012) study did not complete the post-test until one week after the participants had completed the FCAT, therefore, it is likely that participants felt a sense of relief and naturally experienced less anxiety than they did prior to the test. Additionally, circumstances could have arisen that would have obscured the impact of the intervention. This delayed measurement was a clear limitation of the study design, and led to a low WoE C score for 'Outcome Measure'.

Outcomes

The effect sizes calculated for all the studies was the standardised mean difference (Cohen's *d*). Cohen's (1988) *d* descriptors were used to describe the effect sizes (see Table 5). Calculated effect sizes are provided in Table 6. Effect sizes for studies with an RCT design were calculated by inputting the means and standard deviations reported within the papers into Campbell's Collaboration online calculator (Wilson, n.d.). As Campbell (2012) used a quasi-experimental design with one group, the effect size was calculated by subtracting the pre-group mean from the post-group mean and dividing by the SD at pre (Becker, 1988).

Campbell (2012) found no significant difference in participants' test anxiety scores after the intervention. It should be noted that the study took place in a

busy conference room and meetings were often interrupted. These issues were not reported in the other studies. A possible reason for this outcome is that the distractions or excessive challenges could have disrupted the flow experience (Nakamura et al., 2009). This indicates that environmental factors such as noise and disruptions should be carefully managed in practice.

In four of the studies (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019; Morrell, 2018), test anxiety was induced using experimental manipulation. In the Campbell (2012) study, researchers did not include tests to induce anxiety as participants were preparing to sit a national exam during the time of the study. The use of a real testing situation increases the external validity of the findings as participants likely would have attached more importance to the test, especially as they could have received rewards or penalties based on their results (Campbell, 2012).

In the remaining studies (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019; Morrell, 2018), participants were told they were going to complete a test as part of their involvement in the study. This can bring into question the validity of the findings, particularly as research has found anxiety levels increased in students prior to final exams, but not mock exams (Lotz & Sparfeldt, 2017). This indicates that any anxiety the participants experienced is not representative of how they may feel in real testing situations where the stakes are higher. Therefore, the outcomes of these studies must be taken with caution as they do not represent how the use of mandala interventions can affect CYP's level of test anxiety in a real-life context.

As noted previously, four studies (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019; Morrell, 2018) received a 'high' WoE C rating for 'Outcome Measures'. Administering the test anxiety measure (STAIC-S) immediately pre and post intervention reduces the risk of extraneous variables obscuring the impact of the intervention. Consequently, any changes to participants' test anxiety can be more confidently attributed to the use of mandala interventions. Additionally, the post-intervention measure in these studies was administered before participants completed the test. Therefore, the final results of these studies likely yielded more valid measures of test anxiety as participants were in an evaluative situation, a defining feature of test anxiety (Zeidner & Mathews, 2005).

Conversely, participants in the Campbell (2012) study completed their pre-intervention measure two weeks prior to the FCAT. They then engaged in 10-15-minute mandala interventions weekly until the fourth week where they completed the post-test, one week after the FCAT. Given the brief nature of mandala interventions and the leng th of time between pre and post, it is possible that participants' test anxiety increased as they approach the test dates and reduced shortly after they had completed the mandala interventions. However, as test anxiety was measured at two broad time-points, it is difficult to confidently ascertain the impact of the intervention. Additionally, there is an increased risk of extraneous variables such as participants' moods and environmental stressors obscuring the impact of the intervention. Therefore, findings of this study must be taken with caution.

Three studies reported a significant decrease in test anxiety following the mandala intervention (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019), and these resulted in negligible to medium effects (d = 0.58, d = -0.06, d = 0.03, respectively). Morrell (2018) reported a non-significant (p = 0.08) yet marginal reduction of anxiety in participants in the intervention group. Although Morrell (2018) found a medium size effect (d = -0.72) and received a medium WoE D rating, the findings may be somewhat limited as a power analysis revealed that the study was underpowered. Similarly, Carsley et al. (2015) also found a medium effect (d = 0.58, 95% CI, 0.02, 1.13) despite reporting a significant decrease in anxiety scores in both groups. Given the wide confidence intervals, and that this study was underpowered which can lead to unreliable findings (Maxwell, 2004), further research with a larger sample is needed to find the true effect. Caution must therefore be taken when interpreting the current findings.

Two studies found negative effect sizes (Carsley & Heath, 2018; Morrell, 2018) as the intervention group had a lower mean post score of anxiety than the control group. Conversely, Carsley et al. (2015) and Carsley & Heath (2019) found that the intervention group had a slightly higher post score of anxiety than the control group, leading to positive effect sizes. This suggests that between mandala colouring and free colouring, one is not substantially more effective than the other.

Table 5

Descriptors for Cohen's d (Cohen, 1992)

Effect Size	Descriptor
0.8	Large
0.5	Medium
0.2	Small

Table 6:

Effect Sizes for Intervention Outcomes

Research	Sample Size	Design	Outcome Measure	Effect size (*d*)	*p*	WoE D
Campbell (2012)	6	Quasi-experimental design	Children's Test Anxiety Scale (CTAS)	0.27 (Small)	> 0.05	1.2 Low
Carsley, Heath & Fajnerova (2015)	52	Randomised Control Trial	State-Trait Anxiety Inventory for Children State form (STAIC-S)	0.58 (Medium)	0.007	2.5 Medium
Morrell (2018)	43	Randomised Control Trial	State-Trait Anxiety Inventory for Children State form (STAIC-S)	-0.72 (Medium)	0.08	2.1 Medium
Carsley & Heath (2018)	193	Randomised Control Trial	State-Trait Anxiety Inventory for Children State form (STAIC-S)	-0.06 (Negligible)	< 0.01	2.4 Medium
Carsley & Heath (2019)	152	Randomised Control Trial	State-Trait Anxiety Inventory for Children State form (STAIC-S)	0.03 (Negligible)	< 0.001	2.4 Medium

Note: the p value is based on the difference between pre and post measures of test anxiety in the intervention group. The effect size (d) represents between groups pre-post change.

Conclusion and Recommendations

Summary

This review examined the effectiveness of mandala interventions at reducing test anxiety in school aged children. Of the five studies included in this review, four received a medium WoE D rating and one received a low rating (Campbell, 2012). All studies measured test anxiety using self-report measures, and involved children partaking in a mandala intervention for 10 – 15 minutes. Three studies found small effects (Campbell, 2012; Carsley & Heath, 2018; Carsley & Heath; 2019) and two found medium effects (Carsley et al., 2015; Morrell, 2018). However, findings of the latter studies must be taken with particular caution as they were both underpowered.

Three studies (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019) included control groups which received comparative treatment; participants engaged in free drawing for 10 – 15 minutes, which may explain why these studies found small to medium effects. Although the lack of structure inherent in a free drawing activity can potentially be challenging (Curry & Kasser, 2005), both the control and intervention groups reported significant reductions in anxiety. However, as the control group interventions were very similar to the mandala intervention, future research may wish to include an inactive control group to allow for clearer comparison.

Strengths and Limitations

To the author's knowledge, there has not been a systematic review focused exclusively on the effectiveness of mandala interventions at reducing test anxiety. Although there has been some research investigating the impact of

mandala colouring (Duong et al., 2018; Noor et al., 2017; Van Der Vennet & Serice, 2012), these have primarily focused on adult participants, and therefore findings cannot be extrapolated to children. This review therefore aimed to provide a comprehensive overview of the effectiveness of mandala interventions in school age children. To ensure that the search was as comprehensive as possible, this review included both peer reviewed articles and unpublished theses. Including peer reviewed articles provides the assurance that the articles have been subject to scientific rigour by trained researchers. Although this level of quality assurance is not something that unpublished theses are subjected to, it was deemed that their inclusion in this review reduced the risk of publication bias, and provided valuable empirical data which assisted in exploring the review question.

A limitation of this review is that the included studies did not report including participants with anxiety disorders, or participants from a range of diverse backgrounds. As such, the findings of this review may not generalise to these populations. Caution must be taken in using the results to draw wide conclusions about how these children might respond to mandala and free colouring interventions.

Experimental manipulation was used to induce test anxiety in four of the studies. Research indicates that pupils experience higher levels of test anxiety in relation to high-stakes assessment than to classroom tests (Segool et al., 2013). As a result, participants likely would have attached less importance to the tests used, leading to lower motivation to perform well (McDonald, 2001). Therefore, further research is needed to explore whether

mandala interventions are effective for students who are preparing to take more formal examinations such as A-Levels.

All studies used self-report methods to measure test anxiety. Although the studies reported high internal consistency of both the CTAS and STAIC, test anxiety of the participants was measured using just one form of self-report after a 15-minute intervention. This led to WoE A penalties for all studies, as the lack of triangulation of data can affect the reliability and validity of the findings (Moon, 2019). Therefore, future research should consider exploring the longer-term effects of mandala interventions, and use multiple measures to monitor impact. This can include physiological measures which can be an objective way of assessing arousal in performance-evaluative situations (Houtveen & de Geus, 2009).

Implications for Future Practice and Research

The findings of this review suggest that mandala interventions prior to low stakes tests can be effective for some children. However, as the studies did not include participants with diagnosed anxiety disorders and those from a range of ethnic backgrounds, further research is needed to explore whether similar findings will be replicated with these populations. Additionally, as this review included participants aged nine to fourteen, future research can explore the impact of mandala interventions in older children who typically experience a greater frequency of testing (McDonald, 2001).

Although only small and medium size effects were found, and the studies predominantly recruited Caucasian school age children, this intervention holds promise as a method to reduce test anxiety in school age children in

the UK. As there appears to be comparative benefits of both mandala colouring and free drawing, educational professionals should consider incorporating these activities in their classrooms. Given the limitations discussed in this review however, staff should also account for individual differences which may impact the effectiveness of the intervention. Finally, EPs may wish to signpost the interventions to parents and carers as a cost-effective and time-efficient method to reduce test anxiety in their children.

References

Barker, C., Pistrang, N., & Elliott, R. (2015). *Research methods in clinical psychology: An introduction for students and practitioners.* John Wiley & Sons.

Barrett, C. A., & NC-BC, H. S. M. I. (2015, December). Adult coloring books: Patterns for stress relief. In *Phi Kappa Phi Forum* (Vol. 95, No. 4, p. 27). National Forum: Phi Kappa Phi Journal.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*(2), 257-278.

Beckwith, P. (2014). Mindfulness and mandalas: Alternative therapeutic techniques for AOD adolescents. *Capital University's Undergraduate Research Journal,* 1-5.

Bluth, K., & Blanton, P. W. (2014). Mindfulness and self-compassion: Exploring pathways to adolescent emotional well-being. *Journal of child and family studies, 23*(7), 1298-1309.

Bradley, R. T., McCraty, R., Atkinson, M., Tomasino, D., Daugherty, A., & Arguelles, L. (2010). Emotion self-regulation, psychophysiological coherence, and test anxiety: results from an experiment using electrophysiological measures. *Applied psychophysiology and biofeedback, 35*(4), 261-283.

Breedlove, S. M., & Watson, N. V. (2013). *Biological psychology: An introduction to behavioral, cognitive, and clinical neuroscience.* Sinauer Associates.

Campbell, A. C. (2012). Reduction of test anxiety by using mandalas: A pilot study (Order No. 1519269). Available from ProQuest Dissertations & Theses Global. (1095541548). Retrieved from https://search.proquest.com/dissertations-theses/reduction-test-anxiety-using-mandalas-pilot-study/docview/1095541548/se-2?accountid=14511

Carsley, D., & Heath, N. L. (2018). Effectiveness of mindfulness-based colouring for test anxiety in adolescents. *School Psychology International, 39*(3), 251-272.

Carsley, D., & Heath, N. L. (2019). Evaluating the effectiveness of a mindfulness coloring activity for test anxiety in children. *The Journal of Educational Research, 112*(2), 143-151.

Carsley, D., Heath, N. L., & Fajnerova, S. (2015). Effectiveness of a classroom mindfulness coloring activity for test anxiety in children. *Journal of Applied School Psychology, 31*(3), 239-255.

Chen, H., Liu, C., Chiou, W. K., & Lin, R. (2019, July). How Flow and Mindfulness Interact with Each Other in Different Types of Mandala Coloring Activities? In *International Conference on Human-Computer Interaction* (pp. 471-486). Springer, Cham.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillside, NJ: Erlbaum

Csikszentmihalyi, M. (1975). Beyond boredom and anxiety. San Francisco, California: Jossey-Bass.

Csikszentmihalyi, M. (2002). Flow: The classic work on how to achieve happiness. Random House.

Curry, N. A., & Kasser, T. (2005). Can coloring mandalas reduce anxiety? *Art Therapy, 22*(2), 81-85.

Department for Education. (2015, January). SEND code of practice: 0 to 25 years. GOV.UK. https://www.gov.uk/government/publications/send-code-of-practice-0-to-25

DfE. (2014). Mental health and behaviour in Schools: Departmental advice for school staff (1st ed.). (June).

Duong, K., Stargell, N. A., & Mauk, G. W. (2018). Effectiveness of coloring mandala designs to reduce anxiety in graduate counseling students. *Journal of Creativity in Mental Health*, *13*(3), 318-330.

Edenfield, T. M., & Saeed, S. A. (2012). An update on mindfulness meditation as a self-help treatment for anxiety and depression. *Psychology Research and Behavior Management*, *5*, 131.

Ergene, T. (2003). Effective interventions on test anxiety reduction. *School Psychology International, 24*, 313–328.

Fridrici, M., & Lohaus, A. (2009). Stress-prevention in secondary schools: online- versus face-to-face-training. *Health Education, 109*(4), 299–313. https://doi.org/10.1108/09654280910970884.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional children, 71*(2), 149-164.

Goldin, P. R., & Gross, J. J. (2010). Effects of mindfulness-based stress reduction (MBSR) on emotion regulation in social anxiety disorder. Emotion, 10(1), 83.

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

Greenberg, M. T., & Harris, A. R. (2012). Nurturing mindfulness in children and youth: Current state of research. *Child Development Perspectives, 6*(2), 161-166.

Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG: an international journal of obstetrics and gynaecology, 125*(13), 1716.

Henderson, P., Mascaro, N., Rosen, D. & Skillern, T. (2007). Empirical study on the healing nature of mandalas. *Psychology of Aesthetics, Creativity and the Arts, 1*(3), 148-154.

Hernández, S. E., Suero, J., Barros, A., González-Mora, J. L., & Rubia, K. (2016). Increased grey matter associated with long-term sahaja yoga meditation: a voxel-based morphometry study. PloS One, 11(3), e0150757

Houtveen, J. H., & de Geus, E. J. (2009). Noninvasive psychophysiological ambulatory recordings: Study design and data analysis strategies. *European Psychologist, 14*(2), 132.

Lotz, C., & Sparfeldt, J. R. (2017). Does test anxiety increase as the exam draws near? – Students' state test anxiety recorded over the course of one semester. *Personality and Individual Differences*, *104*, 397-400.

Lowe, P. A., Lee, S. W., Witteborg, K. M., Prichard, K. W., Luhr, M. E., Cullinan, C. M., … Janik, M. (2008). The test anxiety inventory for children and adolescents (TAICA): Examination of the psychometric properties of a new multidimensional measure of test anxiety among elementary and secondary school students. *Journal of Psychoeducational Assessment, 26*, 215–230.

Mackey, A. P., Finn, A. S., Leonard, J. A., Jacoby-Senghor, D. S., West, M. R., Gabrieli, C. F., & Gabrieli, J. D. (2015). Neuroanatomical correlates of the income-achievement gap. *Psychological Science, 26*(6), 925-933.

Mao, Y., Yang, R., Bonaiuto, M., Ma, J., & Harmat, L. (2020). Can Flow Alleviate Anxiety? The Roles of Academic Self-Efficacy and Self-

Esteem in Building Psychological Sustainability and Resilience. *Sustainability, 12*(7), 2987.

Marques, S. C., Pais-Ribeiro, J. L., & Lopez, S. J. (2011). The role of positive psychology constructs in predicting mental health and academic achievement in children and adolescents: A two-year longitudinal study. *Journal of Happiness Studies, 12*(6), 1049-1062.

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychological Methods, 9*(2), 147.

McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational psychology, 21*(1), 89-101.

Merikangas, K. R., He, J., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., et al. (2010). Lifetime Prevalence of Mental Disorders in U.S. Adolescents: Results from the National Comorbidity Survey Replication-Adolescent Supplement (NCS-A). *Journal of the American Academy of Child and Adolescent Psychiatry, 49*(10), 980– 989

Moon, M. D. (2019). Triangulation: A method to increase validity, reliability, and legitimation in clinical research. *Journal of Emergency Nursing, 45*(1), 103-105

Morrell, M. E. (2018). *Test anxiety reduction through brief mindfulness meditation and mandala coloring: School-based mindfulness interventions for urban youth* (Order No. 10934041). Available from ProQuest Central; ProQuest Dissertations & Theses Global.

(2129613580). Retrieved from https://search.proquest.com/dissertations-theses/test-anxiety-reduction-through-brief-mindfulness/docview/2129613580/se-2?accountid=14511

Nakamura, J., & Csikszentmihalyi, M. (2009). Flow theory and research. *Handbook of positive psychology*, 195-206.

Napoli, M., Krech, P. R., & Holley, L. C. (2005). Mindfulness training for elementary school students: The attention academy. Journal of Applied School Psychology, 21(1), 99-125.

Noor, S. M., Saleem, T., Azmat, J., & Arouj, K. (2017). Mandala-coloring as a therapeutic intervention for anxiety reduction in university students. *Pakistan Armed Forces Medical Journal*, *67*(6), 904-07.

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. J*ournal of Epidemiology & Community Health, 57*(7), 527-529.

Pope, D. (2010). Beyond. *Education Canada, 50*(1), 4-8.

Putwain, D. W. (2008). Deconstructing test anxiety. *Emotional and Behavioural Difficulties, 13*(2), 141-155.

Putwain, D., & Daly, A. L. (2014). Test anxiety prevalence and gender differences in a sample of English secondary school students. *Educational Studies, 40*(5), 554-570.

Rose, S. E., & Lomas, M. H. (2020). The Potential of a Mindfulness-Based Coloring Intervention to Reduce Test Anxiety in Adolescents. *Mind, Brain, and Education, 14*(4), 335-340

Sapp, M., Farrell, W., & Durand, H. (1995). The effects of mathematics, reading, and writing tests in producing worry and emotionality test anxiety with economically and educationally disadvantaged college students. *College Student Journal, 29,* 122-125.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: reactions to tests. *Journal of Personality and Social Psychology*, 46(4), 929.

Segool, N., Carlson, J., Goforth, A., von der Embse, N., & Barterian, J. (2013). Heightened test anxiety among young children: Elementary school students' anxious responses to high-stakes testing. *Psychology in the Schools, 50*, 489–499.

Soares, D., & Woods, K. (2020). An international systematic literature review of test anxiety interventions 2011–2018. *Pastoral Care in Education, 38*(4), 311-334.

Spielberger, C. D., & Gorsuch, R. L. (1973). *State-trait Anxiety Inventory for Children: Sampler Set: Manual, Test Booklet and Scoring Keyl*. Mind Garden.

Sportel, B. E., de Hullu, E., de Jong, P. J., & Nauta, M. H. (2013). Cognitive bias modification versus CBT in reducing adolescent social anxiety: a randomized controlled trial. *PLoS One, 8*(5), e64355.

Stratton, S. J. (2019). Quasi-experimental design (pre-test and post-test studies) in prehospital and disaster research. *Prehospital and disaster medicine, 34*(6), 573-574.

Swanson, S., & Howell, C. (1996). Test anxiety in adolescents with learning disabilities and behavior disorders. Exceptional Children, 62, 389–397. doi: 10.1177/001440299606200501

Taren, A. A., Creswell, J. D., & Gianaros, P. J. (2013). Dispositional mindfulness co-varies with smaller amygdala and caudate volumes in community adults. *PloS One, 8*(5), e64574.

Teasdale, J. D., Segal, Z. V., & Williams, J. M. (1995). How does cognitive therapy prevent depressive relapse in recurrent major depressive disorder: A systematic review and meta-analysis. *Clinical Psychology Review, 33*, 25-39.

Thorley, C. (2016). Supporting secondary schools to play a central role in early intervention mental health services. Institute for Public Policy Research. Retrieved from https://www.ippr.org/files/publications/pdf/education-education-mental-health_summary_May2016.pdf

Van Der Vennet, R., & Serice, S. (2012). Can coloring mandalas reduce anxiety? A replication study. *Art Therapy*, *29*(2), 87-92.

von der Embse, N. P., Kilgus, S. P., Segool, N., & Putwain, D. (2013). Identification and validation of a brief test anxiety screening tool.

International Journal of School & Educational Psychology, 1(4), 246-258.

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders, 227*, 483-493.

Wilson, D. B. (n.d.). Campbell Collaboration Effect Size Calculator. Retrieved January 29, 2021, from https://www.campbellcollaboration.org/research-resources/effect-size-calculator.html

Wren, D. G., & Benson, J. (2004). Measuring test anxiety in children: Scale development and internal construct validation. *Anxiety, Stress & Coping, 17*(3), 227-240.

Zeidner, M., & Matthews, G. E. R. A. L. D. (2005). Evaluation anxiety. *Handbook of Competence and Motivation*, 141-163.

Appendices

Appendix A: List of Excluded Studies

Excluded Research	
Reference	**Stage of exclusion and exclusion criteria**
de Campos, F. V., Antunes, C. F., Damião, E. B. C., Rossato, L. M., & Nascimento, L. C. (2020). Anxiety assessment tools in hospitalized children. *Acta Paulista de Enfermagem, 33*(5), 1-8.	Abstract screening 3 – This study is a review of literature.
Sinha, N. (2017). Gender Differences in Academic Stress and Test Anxiety Among Resilient Adolescents. *Indian Journal of Psychological Science*, *8*(2), 1-11.	Abstract screening 4 – This study does not include colouring or drawing mandala as the intervention.
Rose, S. E., & Lomas, M. H. (2020). The Potential of a Mindfulness-Based Coloring Intervention to Reduce Test Anxiety in Adolescents. *Mind, Brain, and Education*, *14*(4), 335-340.	Full article reading 3 – No empirical data as data collection has not yet started.

Appendix B: Mapping the Field

Authors	Participants	Design and groups	Setting	Description of methodology	Pre and post anxiety measures
Campbell (2012)	Sample size: 6 Ages: 9 to 12 Females: 5 Males: 1 Participants were recruited using convenience sampling. Students' maths teacher referred participants based on their displays of anxious behaviour.	Quasi-experimental design (one group pretest-posttest design)	Private elementary school in Florida.	Participants attended three 20-25-minute meetings where they were provided with a blank piece of paper and created their own mandala drawings. Test anxiety was not induced by the researcher as the study took place prior to the Florida Comprehensive Assessment Test (FCAT)	Children's Test Anxiety Scale (CTAS)
Carsley, Heath & Fajnerova (2015)	Sample size: 52 Mean age: 10.92 *SD* = 0.82 Females: 28 Males: 24 Participants were recruited using convenience sampling; teachers	Randomised control trial Groups: Mandala = 26 Control = 26	Private elementary school in Montreal, Canada.	Researchers conducted the 50-minute study on three separate days for each of the three grades. To induce test anxiety, participants were told that they would be given a spelling test and the results would be shared with their parents. Participants then had 15 minutes to colour	State-Trait Anxiety Inventory for Children State form (STAIC-S)

	distributed consent forms to grade 4, 5 and 6 elementary school students.			structured mandalas or colour on a blank sheet.	
Morrell (2018)	Sample size: 43 Ages: 12 to 14 Mean age: 13 Females: 24 Males: 19 Participants were recruited using convenience sampling; letters about the research were sent to parents of 280 students at two elementary schools.	Randomised control trial Groups: Mindfulness = 14 Mandala = 15 Control = 14	Middle school in Chicago, Illinois.	To induce test anxiety, participants were told they will be taking a test, and that they would be ranked on their performance. Participants then had 15 minutes to partake in their activity; participants in the mandala group coloured in a structured mandala, those in the mindfulness group listened to a guided mindfulness script, and those in the control group read short stories. .	State-Trait Anxiety Inventory for Children State form (STAIC-S)
Carsley & Heath (2018)	Sample size: 193 Mean age: 13.49 *SD* = 0.5 Females: 109 Males: 84 Participants were grade 8 students recruited using convenience	Randomised control trial Groups: Mandala = 97 Control = 96	Two elementary schools in Montreal, Canada.	Participants were told that they would be completing a test, and results would be shared with their parents. After completing the pre-intervention measures, participants in the intervention group spent 15 minutes colouring structured mandalas while those in the control group engaged in free drawing.	State-Trait Anxiety Inventory-State Anxiety Scale (STAIC-S) Child and Adolescent Mindfulness

	sampling; letters were sent to parents of students at two elementary schools.				Measure (CAMM)
Carsley & Heath (2019)	Sample size: 152 Mean age = 10.38 *SD* = 0.88 Females: 76 Males: 76 Convenience sampling was used to recruit participants; consent forms were distributed to students of public elementary schools in Canada.	Randomised control trial Groups: Mindful = 76 Control = 76	Central location and pupils' classrooms.	To elicit test anxiety, all participants were told they will be taking a spelling test and their results will be shared with their parents. After completing pre-intervention measures, all participants were provided with coloured pencils and either a copy of a structured mandala if they were in the intervention group, or a blank piece of paper if they were in the control group. Both groups engaged in their activities for 15 minutes.	State-Trait Anxiety Inventory for Children State form (STAIC-S) Child and Adolescent Mindfulness Measure (CAMM) Mindful Attention Awareness Scale state version (state MAAS)

Appendix C: Weight of Evidence (WoE)

WoE A: Methodological Quality

This section considers the methodological quality of the studies included in this review. The WoE A judgement was made using an adapted version of the Gersten et al. (2005) protocol (See Appendix D for adaptations and rationale).

The tables below show the criteria used to establish an overall WoE A judgement for the studies. The criteria are based on recommendations from Gersten et al. (2005) protocol which states that for a high rating, the study needs to meet all but one essential criteria, and demonstrate meeting at least 50% of the desirable criteria. Therefore, for the protocol used for coding RCT studies (Carsley et al., 2015; Carsley & Heath, 2018; Carsley & Heath, 2019; Morrell, 2018), the study needed to meet at least 9 out of 10 essential criteria and 4 out of 8 desirable criteria (see Table 7). For the protocol used to code the study by Campbell (2012) which utilised a quasi-experimental design with no control group, the study needed to meet at least 6 out of 7 essential criteria and 3 out of 6 desirable criteria (see Table 8).

WoE A Criteria for Randomised Control Trial Studies:

WoE Rating	Criteria
3 (High)	1. Study meets at least 9 essential criteria 2. Study meets 4 or more desirable criteria
2 (Medium)	1. Study meets at least 9 essential criteria 2. Study meets between 1 and 4 desirable criteria
1 (Low)	1. Study meets less than 9 essential criteria

Table 7

WoE A Criteria for One Group Pretest-Posttest Quasi Experiments:

WoE Rating	Criteria
3 (High)	1. Study meets at least 6 essential criteria 2. Study meets 3 or more desirable criteria
2 (Medium)	1. Study meets at least 6 essential criteria 2. Study meets between 1 and 3 desirable criteria
1 (Low)	1. Study meets less than 6 essential criteria

Table 8

Summary of WoE A Ratings:

Study	Number of essential criteria	Number of desirable criteria	WoE A
Campbell (2012)	5	4	1 (Low)
Carsley, Heath & Fajnerova (2015)	9	3	2 (Medium)
Morrell (2018)	8	3	1 (Low)
Carsley & Heath (2018)	9	2	2 (Medium)
Carsley & Heath (2019)	9	2	2 (Medium)

Table 9

WoE B: Methodological Relevance to the Review Question

This section considers the methodological quality of each included study and its relevance for answering the current review question.

Woe B Criteria:

Low (1)	Medium (2)	High (3)
Quasi-experimental or non-experimental design	Quasi-experimental design	Randomised control trials
No control/comparison group	Control/comparison group	Control/comparison group

Table 10

Note: The criteria are based on recommendations from Petticrew and Roberts (2003) which identify RCTs as the most appropriate design for exploring research questions related to effectiveness. Therefore, studies utilising a RCT design are awarded the highest WoE B rating.

Summary of WoE B Ratings:

Study	WoE B
Campbell (2012)	1 (Low)
Carsley, Heath & Fajnerova (2015)	3 (High)
Morrell (2018)	3 (High)
Carsley & Heath (2018)	3 (High)
Carsley & Heath (2019)	3 (High)

Table 11

WoE C: Relevance to the Review Question

This section considers the relevance of the focus of each study to the review question. The coding protocol for WoE C is provided in Table 12 below. The mean score for each category was calculated to provide the overall WoE C rating for each study.

Criteria	WoE C Rating	Rationale
a) Location	1 Study was carried out in a country that is not culturally similar to the UK. 2 Study was carried out in a country that is culturally similar to the UK. 3 Study was carried out in the UK.	To increase the generalisability of the findings. Participants in the UK or a location that is culturally similar to the UK are likely to have similar experiences of schooling and tests.
b) Age of Participants	1. The mean age of participants is unspecified. 2. Participants are of school age (5-16) and the mean age is included. 3. Participants are of school age, and the mean age is included with standard deviations.	This review is looking at the effectiveness of mandala colouring activities on school age children; findings from studies which included participants outside of the 5-16 age range may not be generalisable to school age children.
c) Ethnicity of Participants	1. Information about the participants' ethnic backgrounds is not provided. 2. The sample is not comprised of students from a range of ethnic backgrounds. 3. The sample is comprised of students from at least three different ethnic backgrounds	This will increase the generalisability of findings to a wider range of school age children.
d) Intervention description	1. The activity instructions are unclear and/or there is no paperwork capturing	Information regarding the activity instructions and paperwork documenting the nature of the

Criteria	WoE C Rating	Rationale
	the nature of the intervention 2. The activity instructions are clear, but there is no paperwork capturing the nature of the intervention 3. The activity instructions are clearly described, and there is paperwork capturing the nature of the intervention.	intervention allows for replicability of the activity.
e) Intervention Delivery	1. Intervention was delivered by the researchers outside an educational setting 2. Intervention was delivered by researchers in an educational setting. 3. Intervention was delivered by existing staff in an educational setting.	External validity of the findings will be higher in studies where existing school staff have delivered the intervention in an educational setting.
f) Outcome Measure	1. The post intervention measure was administered more than a day after the intervention 2. The post intervention measure was administered one day after the intervention 3. The post intervention measure was administered shortly after the intervention	It is important that test anxiety is measured shortly after the intervention as this reduces the chances of confounding variables such as time affecting the level of test anxiety.

Table 12 *Ratings and Descriptors: 1 = Low, 2 = Medium, 3 = High.*

Summary of WoE C ratings

Study	WoE C criteria						
	Location	Age of Participants	Ethnicity of Participants	Intervention Description	Intervention Delivery	Outcome Measure	Overall WoE C
	(A)	(B)	(C)	(D)	(E)	(F)	
Campbell (2012)	2	1	1	3	2	1	1.7 (Low)
Carsley, Heath & Fajnerova (2015)	2	3	2	3	2	3	2.5 (Medium)
Morrell (2018)	2	2	2	3	2	3	2.3 (Medium)
Carsley & Heath (2018)	2	3	1	2	2	3	2.2 (Medium)
Carsley & Heath (2019)	2	3	2	2	2	3	2.3 (Medium)

Table 13 *Please note: A score of 1 – 1.9 is low, 2 – 2.9 is medium and a score of 3 is high.*

Appendix D: Adaptations to Coding Protocol

The Gersten et al., (2005) coding protocol was adapted to make it more appropriate and tailored for the studies included in this review. A list of the questions which were removed and/or adapted can be found below.

~~Strikethrough~~ is used to indicate the original question and updated questions, along with questions that the author creator, are in bold. Comments and the rationale for changes can be found in square brackets.

Essential Quality Indicators

- ~~Was sufficient information provided to determine/confirm whether the participants experienced demonstrated the disability(ies) or difficulties presented?~~ **Was the sample population appropriate for the research question?**

  [This review did not focus on reducing test anxiety in a targeted population. Therefore, this updated question aims to seek whether the study met the inclusion criterion relating to including participants who were of school age]

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

  [*This question was removed when coding the study by Campbell (2012) as it used a one group pretest-posttest design. However, for the remaining four RCT studies, this question was kept in.]*

- ~~Was sufficient information given characterizing the interventionists teachers provided? Did it indicate whether they were comparable across conditions?~~

  [This question was removed as it was deemed inappropriate considering the mandala interventions can be implemented without the need for additional training]

- **Was there a clear method to induce and assess test anxiety?**

  [As test anxiety is context-specific, this question was included to ascertain whether test anxiety was likely experienced by participants during the experiment]

- ~~Was the fidelity of implementation described and assessed~~ **Was the delivery of the intervention clearly described?**

  [As mindful colouring is typically a short mindfulness-based task, this question elicits whether the methods can be easily implemented by staff in a school setting]

- ~~Was the nature of services provided in comparison conditions described?~~

  [*This question was removed when coding the study by Campbell (2012) as it used a one group pretest-posttest design. However, for the remaining four RCT studies, this question was kept in.]*

- ~~*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%?*~~ **Is overall attrition less than 30%?**

  [The last question was deemed the most appropriate for these studies considering the brief nature of the tasks]

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

  [*This question was removed when coding the study by Campbell (2012) as it used a one group pretest-posttest design. However, for the remaining four RCT studies, this question was kept in.]*

- ~~Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?~~

  [As test anxiety is related to testing situations (von der Embse et al. (2018), it was deemed that this question was not appropriate. As test anxiety is likely to have naturally reduced after the testing experience, the information gathered from a repeated measure beyond the immediate post-test may not necessarily reflect the impact of intervention.]

- ~~Was any documentation of the nature of instruction or series provided in comparison conditions?~~

  [*This question was removed when coding the study by Campbell (2012) as it used a one group pretest-posttest design. However, for the remaining four RCT studies, this question was kept in.]*

- ~~Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?~~ **Did the research report include actual audio or videotape excerpts or examples of paperwork that capture the nature of the intervention?**
  [Examples of paperwork were deemed appropriate ways to document the nature of the mandala interventions]

Appendix E: Examples of Completed Coding Protocols

Study: Campbell, A. C. (2012). *Reduction of test anxiety by using mandalas: A pilot study* (Order No. 1519269). Available from ProQuest Dissertations & Theses Global.

Essential Quality Indicators

Quality Indicator for Describing Participants

Was the sample population appropriate for the research question(s)?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

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

Was there a clear method to induce and assess test anxiety?

☒Yes – the study took place 2 weeks prior to the Florida Comprehensive Assessment Test (FCAT)

☐No

☐N/A

☐Unknown/Unable to Code

Was the intervention clearly described and specified?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

Was the delivery of the intervention clearly described?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

Quality Indicator for Outcome Measures

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

☐Yes

☒No

☐N/A

☐Unknown/Unable to Code

Quality Indicators for Data Analysis

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?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

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

☐Yes

☒No

☐N/A

☐Unknown/Unable to Code

Desirable Quality Indicators

Is overall attrition less than 30%?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

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

☐Yes

☒No

☐N/A

☐Unknown/Unable to Code

Was evidence of the criterion-related validity and construct validity of the measures provided?

☐Yes

☒No

☐N/A

☐Unknown/Unable to Code

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

☐N/A

☐Unknown/Unable to Code

Did the research report include actual audio or videotape excerpts or examples of work that capture the nature of the intervention?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

Were results presented in a clear, coherent fashion?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

Total number of essential quality indicators: 5 out of 7

Total number of desirable quality indicators: 4 out of 6

Overall Rating of Evidence: 3☐ 2☐ 1☒ 0☐

Study: Carsley, D., Heath, N. L., & Fajnerova, S. (2015). Effectiveness of a classroom mindfulness coloring activity for test anxiety in children. *Journal of Applied School Psychology, 31*(3), 239-255.

Essential Quality Indicators

Quality Indicators for Describing Participants

Was the sample population appropriate for the research question(s)?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

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

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

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

Was there a clear method to induce and assess test anxiety?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

Was the intervention clearly described and specified?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

Was the delivery of the intervention clearly described?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

Was the nature of services provided in comparison conditions described?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

Quality Indicators for Outcome Measures

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

☐Yes

☒No

☐N/A

☐Unknown/Unable to Code

Were outcomes for capturing the interventions effect measured at the appropriate times?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

Quality Indicators for Data Analysis

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?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

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

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

Desirable Quality Indicators

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%?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

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

☐Yes

☒No – only internal consistency reliability was reported.

☐N/A

☐Unknown/Unable to Code

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

☐Yes

☐No

☒N/A

☐Unknown/Unable to Code

Was evidence of the criterion-related validity and construct validity of the measures provided?

☐Yes

☒No

☐N/A

☐Unknown/Unable to Code

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

☐N/A

☒Unknown/Unable to Code

Was any documentation of the nature of instruction or series provided in comparison conditions?

☐Yes

☒No

☐N/A

☐Unknown/Unable to Code

Did the research report include actual audio or videotape excerpts or examples of paperwork that capture the nature of the intervention?

☒Yes – an example of the structured mandala was provided.

☐No

☐N/A

☐Unknown/Unable to Code

Were results presented in a clear, coherent fashion?

☒Yes

☐No

☐N/A

☐Unknown/Unable to Code

Total number of essential quality indicators: 9 out of 10

Total number of desirable quality indicators: 3 out of 8

Overall Rating of Evidence: 3☐ 2☒ 1☐ 0☐