Case Study 1: An Evidence-Based Practice Review Report

Theme: School Based Interventions for Learning

How effective are Acceptance and Commitment Therapy based interventions delivered in school settings at reducing psychological distress in children and young people?

Summary

Acceptance and Commitment Therapy (ACT) is a form of psychotherapy. It uses mindfulness and acceptance techniques to try to help people cope with difficult feelings and take action according to their values (Hayes, Luoma, Bond, Masuda, & Lillis, 2006). It typically involves talk therapy and the use of metaphors, stories and mindfulness exercises. ACT has been used recently in secondary schools (e.g. Burckhardt, Manicavasagar, Batterham, & Hadzi-Pavlovic, 2016; Livheim et al., 2015) with the aim of reducing psychological distress (including depression, anxiety and stress) in young people.

This systematic literature review aims to investigate the evidence base for ACT-based interventions delivered in school settings. It will examine if they are effective at reducing psychological distress in children and young people. The main conclusion of the review is that there is not sufficient evidence for the effectiveness of these interventions. Reasons for this include poor methodological quality of studies and lack of replication of significant findings across studies.

Introduction

a. The need for therapeutic work in schools

There is increased concern over internalising disorders such as depression and anxiety in adolescents (Bor, Dean, & Najman, 2014). In the UK, the most recent Office of National Statistics Report found that one in ten children and young people had a diagnosable mental health disorder (Green et al., 2005). It is unknown whether there is now increased prevalence or better diagnosis and greater awareness. However, a recent report on child and adolescent mental health (Frith, 2016) found demand for services has increased in recent years and service users found the current level of access insufficient. Schools have been seen as appropriate locations for delivering psychological therapies since they may be more accessible than clinics, especially for children from disadvantaged families (Wolpert et al., 2011).

b. Acceptance and Commitment Therapy- Psychological Basis

Acceptance and Commitment Therapy (ACT) is a form of psychotherapy that attempts to change how a person views their thoughts and feelings. This is in contrast to challenging those thoughts directly as in Cognitive Behavioural Therapy (CBT) (Hayes, Strosahl, & Wilson, 1999). ACT is a transdiagnostic approach (Swain, Hancock, Dixon, & Bowman, 2015), meaning it was designed to alleviate a range of psychological conditions as well as chronic pain. ACT does not aim to eliminate psychological distress but helps people to choose to live how they want to despite difficult feelings. However, this is expected to lead to a reduction in psychological distress as people learn to accept their thoughts and feelings, seeing them as constructs rather than objective facts. ACT may currently be of interest due to the recent surge in popularity of mindfulness-based interventions in schools to enhance

mental health (Kallapiran, Koo, Kirubakaran, & Hancock, 2015). There are concerns that ACT and related therapies (e.g. Mindfulness Based Cognitive Therapy), classed as 'third wave', are not grounded in sufficient empirical evidence (Bass, van Nevel, & Swart, 2014). 'Third wave' therapies tend to focus on the function rather than the content of thoughts and behaviours as well as a person's relationship to their thoughts and behaviours, which is referred to as 'context'. (Hayes, 2004). Third wave therapies often include elements of mindfulness, acceptance and spirituality. The 'first wave' of behaviour therapies focused on behaviourist principles such as classical operant conditioning and the 'second wave' focused on cognition (Hayes, 2004). ACT researchers argue that ACT is not less effective than traditional CBT but that it currently lacks good quality research (Swain et al., 2015).

According to its founder, Steven Hayes, ACT is based on Skinner's radical behaviourism whereby thoughts, feelings and language are regarded as internal 'behaviours' (Hayes et al., 2006). ACT targets the context and function of problematic thoughts rather than the thoughts themselves (Hayes, 2004) and is thus sometimes referred to as a contextual behaviour therapy. Psychological distress is thought to arise through normal, rather than pathological, human thought processes (Harris, 2006). Through interacting with their environments, people learn problem-solving processes that allow them to avoid or remove external threats. They then apply these processes to internal threats such as distressing thoughts and feelings. However, trying to control or avoid uncomfortable emotions and unhelpful thoughts can paradoxically exacerbate and strengthen them as the person unwittingly reinforces them, increasing their functional importance (Hayes et al., 2006). ACT provides an alternative way of approaching psychological distress through accepting

it and choosing to live life according to values rather than in order to avoid difficult thoughts and feelings.

The ACT literature does not tend to give precise definitions of psychological distress, preferring to regard ACT as a general approach to the problem of human suffering (Hayes, Strosahl & Wilson, 1999). Hayes, Strosahl & Wilson (1999) use terms such as 'psychological distress', 'human suffering', 'psychopathology', 'psychological problems' and 'psychological suffering' interchangeably. For the purpose of this review, the term 'psychological distress' was adopted. This can be defined as 'mostly depression and anxiety symptoms that are indicative of a more or less intense feeling of emotional ill-being' (Drapeau et al., 2010, p1).

Whilst depression and anxiety symptoms, rather than stress, are usually cited as the main features of psychological distress, there are theoretical and practical grounds for also including stress in the definition. Stress can be both positive (eustress) or negative (distress), as explored in the work of Selye (1974). However, scales used in psychological research tend to capture negative stress or 'distress' (Wheaton & Montazer, 2012), conceptualising it as a negative emotional state (Lovibond & Lovibond, 1995). Research has found that stress, as measured on some scales, may not be distinguishable from anxiety in young adolescents aged 11-14 years (Szabó & Lovibond, 2006). Stress also contributes to feelings of 'emotional ill-being' (Drapeau et al., 2010). High levels of negative stress are likely to cause suffering in the school age population and thus be of practical significance. Using the term 'psychological distress' in this review rather than a specific clinical term is more consistent with the theory behind ACT, which rather than a diagnostic approach, takes a dimensional view of mental health as emotional suffering (Waters, 2012). ACT employs six core

components to attempt to increase psychological flexibility and change the way the person interacts with their thoughts and feelings:

- Cognitive defusion (recognising thoughts as constructs, not objective facts)
- Acceptance
- Contact with the present moment (mindfulness)
- Observing the self
- Values
- Committed action

A typical ACT session includes experiential exercises where therapists or facilitators use stories, metaphors, mindfulness and language games to help people gain a different perspective on their thoughts and emotions (Harris, 2006). ACT was designed as an individual therapy but has been delivered as group psychotherapeutic or psychoeducational sessions in schools (Burckhardt, Manicavasagar, Batterham, & Hadzi-Pavlovic, 2016; Livheim et al., 2015). In one school, ACT was used as a basis for a life skills curriculum (Dixon, 2013). Adaptations for adolescents typically include using concrete props to illustrate metaphors, writing down thoughts in thought bubbles and using art and role-play (Halliburton & Cooper, 2014).

A recent review found emerging evidence that ACT can be effective in reducing symptoms of pain and psychological distress in children (Swain et al., 2015). However, most studies reviewed took place in clinical settings rather than schools. ACT has been found to be more effective than no treatment or placebo (Halliburton & Cooper, 2014) and one study found ACT outperformed traditional CBT in treating adolescent inpatients with depression (Hayes, Boyd, & Sewell, 2011).

c. Relevance to Educational Psychology Practice and Rationale for Review

Educational Psychologists (EPs) work with children and young people with a range of Special Educational Needs and Disabilities (SEND). This now explicitly includes 'Social, Emotional and Mental Health' needs, since the publication of the new SEND Code of Practice (Department for Education; Department of Health, 2014). A review of the roles and functions of EPs (Farrell et al., 2006) found service users wanted more therapy and intervention support. Mackay (2007) argues that EPs are particularly well-placed to deliver psychological therapies in schools due to their high level of training and position in the education system. There is ongoing discussion about how much EPs should engage in therapeutic work in schools (Wade, 2016). However, if they do this, they must make sure they are aware of what constitutes evidence-based practice (Dunsmuir & Hardy, 2016). This means having knowledge of different psychotherapeutic practices supported by high quality published studies.

ACT may be more developmentally appropriate for adolescents than younger children, requiring the use of abstract reasoning skills (Halliburton & Cooper, 2014). Mental health disorders often occur for the first time in adolescence (Murphey, Barry, & Vaughn, 2013), making access to services particularly important. There is very little evidence of the effectiveness of ACT with children under 11 (Swain et al., 2015).

Therefore, the question addressed in this review is:

How effective are Acceptance and Commitment Therapy based interventions delivered in school settings at reducing psychological distress in children and young people?

Critical Review of the Evidence

a. Literature Search and Screening Strategy

Searches were carried out in January 2017 in peer-reviewed journals using electronic databases PsychInfo, Medline and ERIC. These databases were chosen to give access to articles published in journals relating to health, psychology and education. The search included the following terms:

"acceptance and commitment therapy"

AND

child* or adolescen* or teenag* or school or young

The database search produced 390 text results. Titles and abstracts were screened according to inclusion criteria (Table 1) and 376 articles either did not meet criteria and were excluded or were found to be duplicates. Fourteen remaining articles were read in full and 10 were excluded with rationale (see Appendix A). An ancestral search was carried out on remaining articles and one further paper was identified. A graphical representation of this can be seen in Figure 1. One paper included two studies using different samples (Livheim et al., 2015). These two studies will be referred as Livheim et al., 2015a and Livheim et al., 2015b.

b. Weight of Evidence

The studies included in the review were evaluated according to Gough's Weight of Evidence (WoE) framework (Gough, 2007). Each study was judged on methodological quality (WoE A), methodological relevance (WoE B) and topic relevance (WoE C). The scores for each of the three categories were summed and

divided by three to give an overall Weight of Evidence rating for the study (WoE D). Table 3 summarises the WoE scores given to each study in this review. In order to assess WoE A, two different protocols were used since the review includes studies using group designs and one with a small-N design. Protocols adapted from Kratochwill et al. (2003) and Horner et al. (2005) were used to evaluate WoE A for group designs and the small-N design respectively.

WoEs were assigned descriptions of ‘high’ if between 2.4 and 3.0, ‘medium’ if between 1.6 and 2.3 and ‘low’ if 1.5 or under.

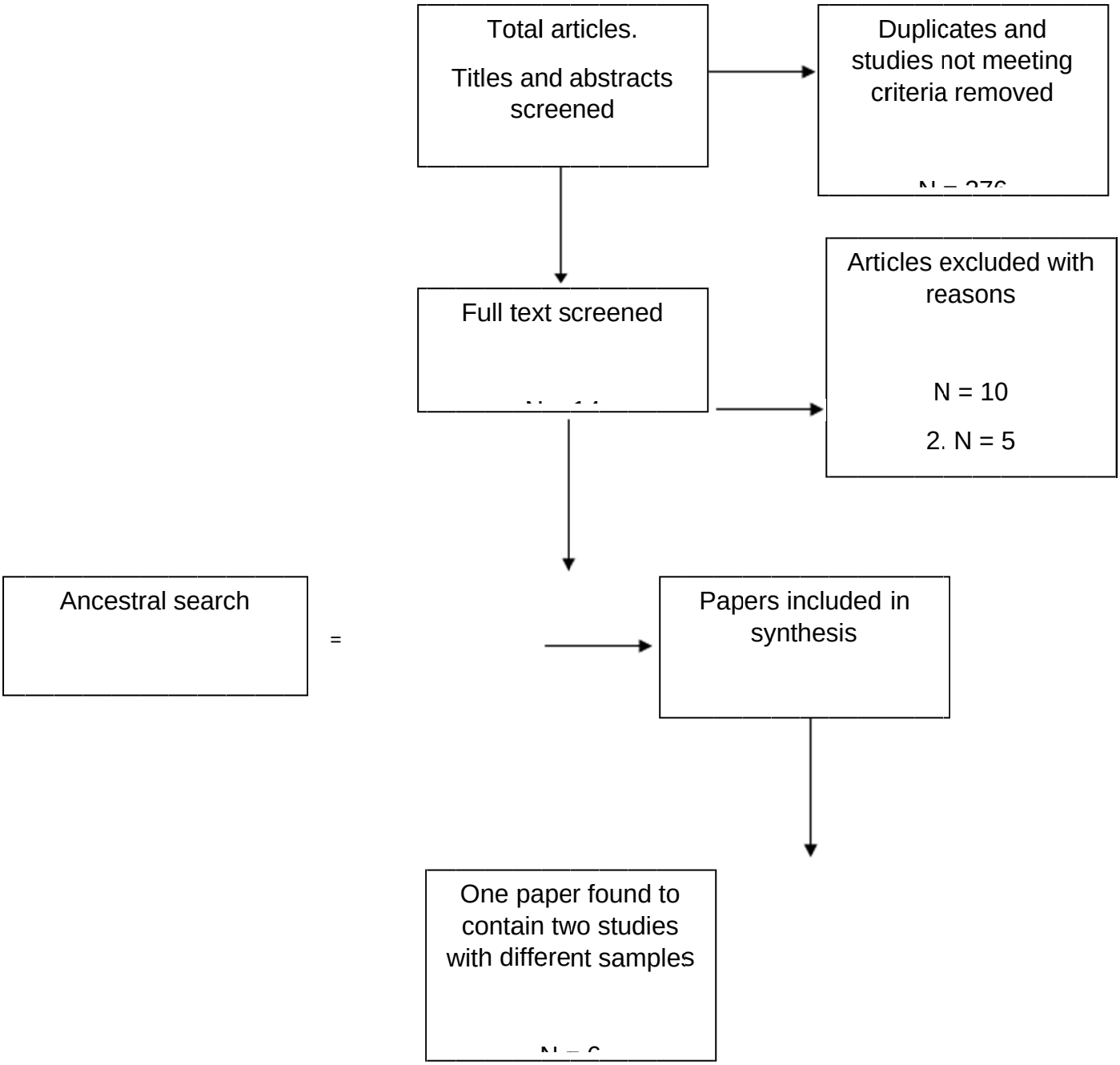

Figure 1. Literature search strategy

Table 1
Inclusion and Exclusion Criteria

Criteria	Inclusion	Exclusion Criteria	Rationale
1. Participants	Study participants are children and young people who are attending school. This will mostly include children and young people aged 11-18, although children and young people aged up to 21 will be included if still attending a school setting (e.g. some special schools provide education for older students).	Participants are younger than 11. Participants are not attending a school setting.	There is very little evidence for the effectiveness of ACT with children under 11 (Swain et al., 2015). EPs work with children and young people up to the age of 25 (Department for Education; Department of Health, 2014) but EPs mostly work in school settings and the implementation of psychological therapies in school settings will be of most interest to them.
2. Setting	Study is set in a school.	Study is in home setting, clinic, university college or preschool	ACT has found to be effective for children and young people when delivered in clinics but no review has yet looked at whether it is effective when

Criteria	Inclusion	Exclusion Criteria	Rationale
			delivered in schools.
3. Intervention	Intervention is based on ACT.	Intervention is not based on ACT	Review aims to investigate effectiveness of ACT.
4. Type of study	Study is empirical and primary in nature and uses an experimental or quasi-experimental design.	Review paper or observational, correlational or qualitative study.	Review aims to collate findings of experimental studies to determine the effectiveness of the intervention.
5. Outcome variables	Study has outcome variables measuring psychological distress (depression, anxiety and stress).	Study does not have outcome variable measuring psychological distress.	Psychological distress was chosen as an outcome to investigate due to current concern about mental health in the school-aged population. Since ACT is designed as a transdiagnostic approach, it is hypothesised to be effective at reducing psychological

Criteria	Inclusion	Exclusion Criteria	Rationale
			distress (depression, anxiety and stress).
6. Date	Date of report- reported after 1999	Reported in 1998 or previous	First major book setting out ACT in its current form was published in 1999.
7. Language	Language of report- English	Language of report is not English	Author does not have resources to translate reports
8. Type of publication	Study must be in a peer-reviewed journal or have been submitted for peer review	Has not been submitted for peer review or published in peer reviewed journal	Peer review acts as primary quality control for studies

Table 2
Full References of Included Studies

Burckhardt, R., Manicavasagar, V., Batterham, P. J., & Hadzi-Pavlovic, D. (2016). A randomized controlled trial of strong minds: A school-based mental health program combining acceptance and commitment therapy and positive psychology. *Journal of School Psychology*, *57*, 41–52. http://doi.org/10.1016/j.jsp.2016.05.008
Livheim, F., Hayes, L., Ghaderi, A., Magnusdottir, T., Högfeldt, A., Rowse, J., Tengström, A. (2015). The Effectiveness of Acceptance and Commitment Therapy for Adolescent Mental Health: Swedish and Australian Pilot Outcomes. *Journal of Child and Family Studies*, 1–15. http://doi.org/10.1007/s10826-014-9912-9
Murrell, A. R., Steinberg, D. S., Connally, M. L., Hulsey, T., & Hogan, E. (2015). Acting Out to ACTing On: A Preliminary Investigation in Youth with ADHD and Co-morbid Disorders. *Journal of Child and Family Studies*, *24*(7), 2174–2181. http://doi.org/10.1007/s10826-014-0020-7
Pahnke, J., Lundgren, T., Hursti, T., & Hirvikoski, T. (2014). Outcomes of an acceptance and commitment therapy-based skills training group for students with high-functioning autism spectrum disorder: A quasi-experimental pilot study. *Autism : The International Journal of Research and Practice*, *18*(8), 953–964. http://doi.org/10.1177/1362361313501091
Theodore-Oklota, C., Orsillo, S. M., Lee, J. K., & Vernig, P. M. (2015). A pilot of an acceptance-based risk reduction program for relational aggression for adolescents. *Journal of Contextual Behavioral Science*, *3*(2), 109–116. http://doi.org/10.1016/j.jcbs.2014.03.001

Table 3
Weight of Evidence of Included Studies

Authors	Methodological Quality WOE A	Methodological Relevance WOE B	Relevance to Review Question WOE C	Overall Weight of Evidence WOE D
Burckhardt, Manicavasagar, Batterham, & Hadzi-Pavlovic, 2016	1.6 (Medium)	2.5 (High)	2.5 (High)	2.2 (Medium)
Livheim et al., 2015 Study A (Australian sample)	1.2 (Low)	1.5 (Low)	2 (Medium)	1.6 (Medium)
Livheim et al., 2015 Study B (Swedish sample)	1.2 (Low)	2 (Medium)	2.5 (High)	1.9 (Medium)
Murrell, Steinberg, Connally, Hulsey, & Hogan, 2015	0.9 (Low)	1 (Low)	1.8 (Medium)	1.2 (Low)
Pahnke, Lundgren, Hursti, & Hirvikoski, 2014	0.8 (Low)	1.5 (Low)	2 (Medium)	1.4 (Low)
Theodore-Oklota, Orsillo, Lee, & Vernig, 2014	1.2 (Low)	1.5 (Low)	1.5 (Low)	1.4 (Low)

c. Participants

In total, 612 participants took part in studies included in this review, aged from 11 to 21. A range of needs was present with considerable heterogeneity across studies. One study did not screen for need (Theodore-Oklota et al., 2014). One study (Burckhardt et al., 2016), conducted separate analyses of data of participants with

depressive or anxious symptoms. One study used referrals from school counsellors. One study (Livheim et al., 2015b) screened for participants. Two studies recruited students based on existing conditions, Autism Spectrum Disorder (ASD) (Pahnke et al., 2014) and Attention Deficit Hyperactivity Disorder (ADHD) (Murrell et al., 2015). This is likely to have impacted on outcomes. Presenting problems of the sample were assessed in WoE C with lower ratings given to studies whose analyses were not clearly of participants starting the study with psychological distress.

Studies took place in a variety of settings in the USA, Sweden and Australia. These are OECD countries with education systems broadly similar to the UK. Therefore, country was not thought to impact highly on WoE for this research question. Two took place in special schools (Murrell et al., 2015; Pahnke et al., 2014), limiting their generalisability to other special schools. Four studies took place in mainstream schools, limiting their generalisability to other mainstream schools. Two of these mainstream schools were particularly affluent environments (Burckhardt et al., 2016; Livheim et al., 2015b). Again, this limits their generalisability to a particularly priviledged student population. One study (Livheim et al., 2015a) included five different schools, one of which was an alternative school for students unable to cope with mainstream educational settings and therefore may have higher external validity since the participants were drawn from multiple settings. Setting was taken into account for WoE C.

Both sexes were represented in the overall sample. Four studies included considerably higher numbers of either boys (Pahnke et al., 2014; Burckhardt et al.,

2016) or girls (Livheim et al., 2015, both studies). This limits the generalisability of these studies to the sex that was studied and was taken into account in WoE C.

Three group studies were underpowered to detect even a large effect (Livheim et al., 2015a and b; Pahnke et al., 2014). Burckhardt et al., (2016) had only enough power to detect a large effect and Theodore-Oklota et al., (2014) had enough power to detect either medium or large effects. Cohen (1992) suggests 26 per group for a large effect, 64 per group for a medium effect and 393 for a small effect when significance level is .05 and power is .08. This could partly explain why findings were inconsistent (Maxwell, 2004). This was reflected when judging WoE A and led to five out of the six studies in this review receiving a 'low' rating in this category.

d. Randomisation

Only one study was fully randomised (Livheim et al., 2015b) and obtained the highest rating on WoE B since randomisation makes groups more likely to be equivalent, allowing researchers to draw more valid conclusions about the true effect of the intervention on the outcome (Barker, Pistrang, & Elliott, 2016). Three studies randomised whole classes to conditions (Burckhardt et al., 2016; Theodore-Oklota et al., 2014; Pahnke et al., 2014). One study randomised female participants but allocated all male participants to the intervention (Livheim et al., 2015a). Murrell et al. (2015) employed a small N design, therefore receiving a low score on WoE B due to substantial threats to internal consistency for which authors did not compensate.

e. Intervention Content and Fidelity

All interventions were based on ACT but their content varied, as did their fidelity. Livheim et al. (2015a and b) used the same manualized 'ACT Experiential Adolescent Group' programme, available online from the Association of Contextual Behavioural Science (https://contextualscience.org/). This programme involves all six components of ACT as well as art and role-play. The Australian study (Livheim et al., 2015a) reported using this manualized version. The Swedish study (Livheim et al., 2015b) translated the programme into Swedish and adapted it to treat stress and to fit into a six week period. Description of adaptations were insufficient for replication and this was reflected in lower WoE A ratings.

Four studies describe ACT-based interventions created by their authors. Pahnke et al., (2014) modified an existing adult ACT protocol although again did not describe adaptations sufficient for replication. Burckhardt et al., (2016) created a new programme, 'Strong Minds', drawing on Positive Psychology and all 6 ACT components. This makes it hard to determine whether ACT components led to participant change. Murrell et al. (2015) based their intervention on a pre-existing protocol, 'ACT for kids', developed by the lead author and a co-author. They describe how they 'carefully noted' adherence to and deviation from 'ACT for kids' but this information is not available in the paper. Theodore-Oklota et al., (2014) did not report using or basing their intervention on a published protocol. Additionally the description of the intervention in the paper does not explicitly include all six ACT components; therefore, this study could only be given a 'low' rating for this aspect of WoE C.

Interventions were delivered by psychology students (Pahnke et al., 2014; Livheim et al., 2015b; Theodore-Oklota et al., 2014), psychologists (Burckhardt et al., 2016; Livheim et al., 2015a) or a 'recognised ACT trainer' (Murrell et al., 2015). One study (Burckhardt et al., 2016) was audiotaped and coded for treatment fidelity on an adherence scale by a psychologist experienced in ACT and therefore scored more highly for WoE A. Where there were comparison groups, these were run by school staff whereas intervention groups were run by psychologists or psychology students. No study counterbalanced change agents so experimenter effects may have influenced outcomes. Facilitator supervision was mentioned explicitly in three studies (Burckhardt et al. 2016; Livheim et al., 2015a; Theodore-Oklota et al., 2014). Supervision was thought to enhance fidelity therefore studies mentioning it explicitly scored higher on WoE A.

f. Comparison Condition

To receive a 'high' rating for comparison condition, studies needed to use an active comparison group also receiving psychotherapeutic support e.g. CBT. No study in this review did this, therefore it compromises findings since many forms of psychotherapy are broadly equivalent in effect (Barker, Pistrang & Elliott, 2016) and it cannot be concluded that the ACT components led to change. One study used a comparison group receiving pastoral care classes of the same length and duration as the ACT group (Burckhardt et al., 2016) and received a 'medium' rating for this criterion. Two involved monitoring or an offer of individual support but not all participants took this up (Livheim et al., 2015 a and b). Two used a waitlist control and one did not use a comparison condition. These obtained a lower score for WoE

B since their design meant that expectancy effects or the Hawthorne effect, rather than the ACT components of the intervention, may have led to change.

g. Measures

Most studies reviewed did not collect data from multiple sources or using multiple methods. When multiple sources were attempted, the measure used was either inappropriate (Pahnke et al., 2014) or staff and parents did not complete questionnaires (Murrell et al., 2015). Therefore, they could not score highly on WoE A. One study collected data on stress using two different scales (Livheim et al., 2015b) and found that participants improved according to one scale but not according to the other.

One study (Pahnke et al., 2014) used inappropriate measures. Authors used the Stress Schedule Survey as a self-report measure although it only has validity and reliability as a parent or teacher-report measure. They also used the Beck Youth Inventories for participants up to age 21 even though they only have validity and reliability for under 18s. This study received a rating of ‘0’ for measurement and its findings must be viewed particularly critically.

h. Outcomes

There is an inconclusive picture of the effectiveness of ACT-based interventions from studies in the present review. Effect sizes for group studies were calculated by subtracting the mean pre-post change in the intervention group from the mean pre-post change in the comparison group and dividing this by the pooled pre-test standard deviation, as recommended by Morris (2008). Effect size for small-N

studies is typically calculated using Percentage of Non-overlapping Data Points (PAND). This was not possible since Murrell et al. (2015) only gave two data points for each participant. Therefore, effect size for this study was calculated by looking at within-participant pre-post change (Becker, 1988). Effect sizes were reported for depression (Table 4), anxiety (Table 5) and stress (Table 6) separately. Effect sizes for psychological distress (Table 7) were reported for studies that used a general measure or also included a composite score. Positive effect sizes mean psychological distress reduced more in the intervention group than the comparison group whereas a negative number means psychological distress reduced more in the comparison group. Effect sizes were calculated even for non-significant findings to investigate whether Type 2 errors were made due to studies being underpowered. Effects were interpreted according to Cohen's criteria (1992) of 0.2 as small, 0.5 as medium and over 0.8 as large. Effects below 0.2 were regarded as negligible.

One study found a medium effect of ACT-based interventions on reducing depression (Burckhardt et al., 2016) and another found a large effect (Livheim et al., 2015a). Two studies found nil effects for this outcome (Livheim et al., 2015b; Pahnke et al., 2014).

Three studies measured anxiety pre and post intervention and all found nil effect on their group of participants as a whole (Burckhardt et al., 2016; Livheim et al., 2015b, Pahnke et al., 2014). However, the Livheim et al. (2015b) study was underpowered and may have produced a Type 2 error due to too few participants (see Table 5). When Burckhardt et al. (2016) reran their analysis, separating participants by year group, they found a medium effect on anxiety in the Year 10 participants only.

Two studies found large effects of ACT-based intervention on stress (Burckhardt et al., 2016; Livheim et al., 2015b) and one study found a medium effect (Pahnke et al., 2014). However, Livheim et al. (2015b) and Pahnke et al. (2014) both used two measures to investigate stress and both found nil effects on the alternative measure. This calls their findings into question.

One study found a medium effect of ACT-based intervention on psychological distress taken as a whole (Burckhardt et al., 2016). Two (Theodore-Oklota et al., 2014; Pahnke et al. 2014) found nil effect of ACT-based intervention on overall psychological distress at post-intervention test. Murrell et al. (2015) found reliable change for five out of seven participants but three of these changed in the undesired direction. Effect size was large, suggesting that psychological distress increased in this group. However, methodology was poor and there were too few participants to trust the results of this analysis.

i. Educational/Clinical Significance

One study (Burckhardt et al., 2016) classified levels of psychological distress according to clinical categories. They found clinical differences on anxiety (for Year 10 only) and stress (for both year groups). Therefore, this study obtained a higher score for WoE A.

j. Follow Up

One study (Pahnke et al., 2014) did a follow up, re-administering measures two months after intervention completion. To avoid multiple comparison issues and Type

1 error, they did not calculate differences between post and follow-up scores separately, presenting figures with results of ANOVAs examining data at times 1, 2 and 3 in a single analysis. Trends show effects of the intervention on stress (Stress Schedule Survey) and psychological distress overall (Beck Youth Inventories total score) becoming significant only at follow-up. Since effects were not calculated, follow-up was not included in WoE ratings.

Table 4
Effect Sizes for Depression

Authors	Measures	ACT			Control			Pretest-Posttest Effect Size (ACT vs control)	p	Effect Size Descriptor	Total WoE
		N	Pre M (SD)	Post M (SD)	N	Pre M (SD)	Post M (SD)				
Burckhardt et al. 2016	Depression subscale of DAS-S 21 (Depression, Anxiety and Stress Scale)	30	25.83 (8.46)	20.17 (9.73)	33	24.30 (7.72)	23.00 (11.69)	0.54	.04*	Moderate	Medium
Livheim et al., 2015a	Reynolds Adolescent Depression Scale	32	66.95 (15.39)	61.57 (21.50)	19	58.54 (15.78)	65.48 (22.36)	0.79	.008**	Large	Medium
Livheim et al., 2015b	Depression subscale of DAS-S 21 (Depression, Anxiety and Stress Scale)	15	7.80 (3.29)	7.98 (4.80)	17	4.40 (3.50)	4.18 (4.82)	-0.11 (control group improved, intervention group became more depressed)	.742	Negligible/no effect	Medium

Authors	Measures	ACT			Control			Pretest-Posttest Effect Size (ACT vs control)	p	Effect Size Descriptor	Total WoE
		N	Pre M (SD)	Post M (SD)	N	Pre M (SD)	Post M (SD)				
Pahnke et al., 2014	Depression subscale of Beck Youth Inventories	15	18.5 (10.84)	14.5 (10.36)	13	14.4 (9.64)	12.25 (9.64)	0.18	.088	Negligible/no effect	Low

* = $p < 0.05$, ** = $p < 0.01$

Table 5
Effect Sizes for Anxiety

Authors	Measures	ACT			Control			Pretest-Posttest Effect Size (ACT vs control)	p	Effect Size Descriptor	WoE
		N	Pre M (SD)	Post M (SD)	N	Pre M (SD)	Post M (SD)				
Burckhardt et al. 2016	Anxiety subscale of DAS-S 21 (Depression, Anxiety and Stress Scale)	30	21.04 (8.93)	19.00 (9.55)	33	20.58 (7.69)	18.64 (11.32)	0.01	.36	Negligible/no effect	Medium
Livheim et al., 2015b	Anxiety subscale of DAS-S 21 (Depression, Anxiety and Stress Scale)	15	7.80 (3.29)	7.98 (4.80)	17	4.40 (3.50)	4.18 (4.82)	0.50	.57	Medium (but non-significant due to lack of power)	Medium
Pahnke et al., 2014	Anxiety subscale of Beck Youth Inventories	15	18.2 (10.75)	17 (12.00)	13	15.5 (9.64)	14.2 (9.55)	0.01	NS (not reported)	Negligible/no effect	Low

Table 6
Effect Sizes for Stress

Authors	Measures	ACT			Control			Pretest-Posttest Effect Size (ACT vs control)	p	Effect Size Descriptor	Total WoE
		N	Pre M (SD)	Post M (SD)	N	Pre M (SD)	Post M (SD)				
Burckhardt et al. 2016	Stress subscale of DAS-S 21 (Depression, Anxiety and Stress Scale)	30	27.00 (6.70)	20.67 (8.38)	33	24.79 (6.03)	23.24 (9.33)	0.75	.006**	Large	Medium
Livheim et al., 2015b	Stress subscale of DAS-S 21 (Depression, Anxiety and Stress Scale)	15	10.40 (4.18)	10.26 (3.60)	17	7.47 (4.45)	7.37 (3.59)	0.009	.971	Negligible/no effect	Medium
	Perceived Stress Scale (PSS)	15	30.00 (5.77)	21.49 (8.44)	17	28.60 (6.14)	26.30 (8.41)	1.04	.009**	Large	Medium

Pahnke et al., 2014	Stress Schedule Survey (Teacher)	15	122 (39.70)	99 (23.24)	13	97.5 (30.42)	93 (36.96)	0.52	NS with multiple comparisons corrections	Medium	Low

Authors	Measures	ACT			Control			Pretest-Posttest Effect Size (ACT vs control)	p	Effect Size Descriptor	Total WoE
		N	Pre M (SD)	Post M (SD)	N	Pre M (SD)	Post M (SD)				
Pahnke et al., 2014	Stress Schedule Survey (Self report)	15	114 (43.57)	102 (41.14)	13	112.5 (58.14)	100 (71.21)	< 0.001	NS (not reported)	Negligible/no effect	Low

* = $p < 0.05$, ** = $p < 0.01$

Table 7
Effect Sizes for Other Measures and Combined Measures of Psychological Distress

Authors	Measures	ACT N	ACT Pre M (SD)	ACT Post M (SD)	Control N	Control Pre M (SD)	Control Post M (SD)	Pretest-Posttest Effect Size	p	Effect Size Descriptor	Total WoE
Burckhardt et al. 2016	Total score on DAS-S 21 (Depression, Anxiety and Stress Scale)	30	73.88 (17.72)	59.83 (23.64)	33	69.67 (14.51)	64.88 (29.78)	0.57	.02*	Medium	Medium
Pahnke et al., 2014	Total score on Beck Youth Inventories	15	51.5 (29.05)	46 (32.44)	13	41 (27.04)	38.5 (27.49)	0.11	NS	Negligible/no effect	Low
	SDQ (Strengths and Difficulties Questionnaire) Emotional Symptoms	15	3.87 (2.97)	3.27 (3.31)	13	2.38 (2.50)	2.62 (1.85)	0.30	NS with corrections for multiple comparisons	Medium (but not significant due to correction for multiple comparisons)	Low
Theodore-Oklota et al., 2014	Child Behaviour Checklist (Youth Self-Report)	105	30.76 (20.04)	31.42 (19.27)	105	31.01 (21.81)	31.92 (22.78)	0.01	NS (not reported)	Negligible/no effect	Low

Authors	Measures	ACT N	ACT Pre M (SD)	ACT Post M (SD)	Control N	Control Pre M (SD)	Control Post M (SD)	Pretest-Posttest Effect Size	p	Effect Size Descriptor	Total WoE
Murrell et al., 2015	Behavior Assessment System for Children (BASC 2)- Emotional Symptoms Index (ESI)	9	48.00 (7.51)	54.29 (17.50)	n/a	n/a	n/a	0.84	NS .31	Large (but in unanticipated and unwanted direction) and non-significant due to study being underpowered.	Low

* = $p < 0.05$

Conclusions and Recommendations

In conclusion, evidence for the effectiveness of ACT-based interventions delivered in school settings at reducing psychological distress (depression, anxiety and stress) in children and young people is limited. This lack of convincing evidence is due to poor methodological quality of the studies reviewed and lack of significant outcomes found in several studies. Five out of six studies reviewed received a rating of 'low' for methodological quality, making it difficult to draw any firm conclusions from their findings.

Studies were heterogeneous in content, setting and sample. Therefore, it is hard to draw conclusions on why some found large effect sizes for some outcomes and others found no effects. It could be a consequence of all but one of the studies being underpowered (Maxwell, 2004). However, the studies that did find significant effects were of better methodological quality (Burckhardt et al., 2016; Livheim et al., 2015a and b) and had some common elements. They included 6 to 16 sessions whereas two studies with worse outcomes (Pahnke et al., 2014; Theodore-Oklota et al., 2014) included only two to three sessions. They were delivered by psychologists rather than by students. They took place in mainstream schools and in two cases the sample had a high socioeconomic status background (Burckhardt et al., 2016; Livheim et al., 2015b). They used measures intended to capture changes in mental health rather than behaviour. This indicates that future research should investigate ACT-based interventions delivered by a psychologist of at least 6 sessions in length, using measures designed to capture changes in mental health. Better quality research is needed, particularly in settings with students attending special schools and from more diverse socioeconomic backgrounds.

The study of the highest methodological quality (Burckhardt et al., 2016) is the only one in this review to provide good evidence of reduction in psychological distress after ACT-based intervention. However, no firm conclusions can be drawn about whether this was due to the ACT components of the intervention. This is because most major forms of therapy are broadly equivalent in effect (Barker et al., 2016). Alternatively, the students who improved may have done so due to expectancy effects or the Hawthorne effect. To understand whether ACT-specific components of the intervention led to change, future research should use a comparison group also receiving psychotherapeutic input e.g. CBT. Despite this, EPs may find this study interesting as a general indication of the potential effectiveness of psychotherapeutic input in schools.

When evaluating psychotherapy, practitioners should be aware that EPs work with adolescents with a range of complex needs. Participants in RCTs and similar designed studies often systematically differ from adolescents seen routinely for psychotherapy, having less severe and fewer comorbid issues (Wolpert, Fugard, & Deighton, 2013). As well as basing practice on evidence, EPs must take into account the local context and presenting problems of the child or young person. ACT may be more appropriate for adolescents experiencing unavoidable psychological distress, for example arising from chronic pain or life-limiting conditions, due to its focus on acceptance. EPs interested in ACT would be advised to proceed with caution due to the current lack of evidence and to carefully evaluate outcomes using Goal Attainment Scaling or Target Monitoring and Evaluation (Dunsmuir, Brown, Iyadurai, & Monsen, 2009). If done rigorously and replicated sufficiently as detailed by

Wolpert, Fugard and Deighton (2013), evaluation could build up a body of practice-based evidence to complement any future RCTs in the field.

Due to the limitations discussed in this review, EPs should be clear that there is currently insufficient evidence to recommend the use of ACT based interventions in schools. Future studies, including evaluation of outcomes by practitioners, should strive for better methodological quality. Future RCTs should ensure they have sufficient power to detect effects and include an active psychotherapy control to identify the effective components of interventions.

References

Barker, C., Pistrang, N., & Elliott, R. (2016). *Research methods in clinical psychology: an introduction for students and practitioners* (3rd ed.). Chichester: Wiley-Blackwell.

Bass, C., van Nevel, J., & Swart, J. (2014). A comparison between dialectical behavior therapy, mode deactivation therapy, cognitive behavioral therapy, and acceptance and commitment therapy in the treatment of adolescents. *International Journal of Behavioral Consultation and Therapy*, *9*(2), 4–8. http://doi.org/10.1037/h0100991

Becker, B. J. (1988). Synthesizing standardized mean change measures. *British Journal of Mathematical and Statistical Psychology*, *41*, 257–278. http://doi.org/10.1111/j.2044-8317.1988.tb00901.x

Bor, W., Dean, A. J., & Najman, J. (2014). Are child and adolescent mental health problems increasing in the 21st century ? A systematic review, *48*(7). http://doi.org/10.1177/0004867414533834

Burckhardt, R., Manicavasagar, V., Batterham, P. J., & Hadzi-Pavlovic, D. (2016). A randomized controlled trial of strong minds: A school-based mental health program combining acceptance and commitment therapy and positive psychology. *Journal of School Psychology*, *57*, 41–52. http://doi.org/10.1016/j.jsp.2016.05.008

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. http://doi.org/10.1037/0033-2909.112.1.155

Department for Education; Department of Health. (2014). *Special educational needs and disability code of practice: 0 to 25 years*. http://doi.org/DFE-00205-2013

Dixon, M. R. (2013). Don’t stop believing: journeys school. *Behavior Analysis in*

Practice, *6*(1), 78–79.

Drapeau, A., Beaulieu-Prévost, D., Marchand, A., Boyer, R., Préville, M., & Kairouz, S. (2010). A life-course and time perspective on the construct validity of psychological distress in women and men. Measurement invariance of the K6 across gender. *BMC Medical Research Methodology*, *10*, 68. http://doi.org/10.1186/1471-2288-10-68

Dunsmuir, S., Brown, E., Iyadurai, S., & Monsen, J. (2009). Evidence-based practice and evaluation: from insight to impact. *Educational Psychology in Practice*, *25*(1), 53–70. http://doi.org/10.1080/02667360802697605

Dunsmuir, S., & Hardy, J. (2016). *Delivering Psychological Therapies in Schools and Communities*. Leicester: British Psychological Society.

Farrell, P., Woods, K., Lewis, S., Rooney, S., Squires, G., & O'Connor, M. (2006). *A review of the functions and contribution of Educational Psychologists in England and Wales in the light of "Every Child Matters: Change for Children". School of Education, University of Manchester*.

Frith, E. (2016). *CentreForum commission on children and young people's mental health: state of the nation*. http://doi.org/10.1080/02650487.1994.11104563

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*(2), 37–41.

Halliburton, A. E., & Cooper, L. D. (2014). Applications and adaptations of Acceptance and Commitment Therapy (ACT) for adolescents. *Journal of Contextual Behavioral Science*, *4*(1), 1–11. http://doi.org/10.1016/j.jcbs.2015.01.002

Harris, R. (2006). Embracing your demons: An overview of acceptance and commitment therapy. *Psychotherapy in Australia*, *12*(4), 2–8.

Hayes, L., Boyd, C. P., & Sewell, J. (2011). Acceptance and Commitment Therapy for the Treatment of Adolescent Depression: A Pilot Study in a Psychiatric Outpatient Setting. *Mindfulness*, *2*(2), 86–94. http://doi.org/10.1007/s12671-011-0046-5

Hayes, L., & Rowse, J. (2008). *Acceptance and Commitment Therapy: Experiential Adolescent Group*.

Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behavior Therapy*, *35*(4), 639–665. http://doi.org/10.1016/S0005-7894(04)80013-3

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and Commitment Therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, *44*(1), 1–25. http://doi.org/10.1016/j.brat.2005.06.006

Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change*. New York: Guilford Press.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*(2), 165–179. http://doi.org/10.1177/001440290507100203

Kallapiran, K., Koo, S., Kirubakaran, R., & Hancock, K. (2015). Review: Effectiveness of mindfulness in improving mental health symptoms of children and adolescents: A meta-analysis. *Child and Adolescent Mental Health*, *20*(4), 182–194. http://doi.org/10.1111/camh.12113

Kratochwill, T., Fuchs, L., Abidin, D., Hughes, J., Batch, G., Keith, T., … Strein, W. (2003). *Task Force on Evidence-Based Interventions in School Psychology.*

Procedural and Coding Manual. Retrieved from http://www.indiana.edu/~ebi/documents/_workingfiles/EBImanual1.pdf

Livheim, F., Hayes, L., Ghaderi, A., Magnusdottir, T., Högfeldt, A., Rowse, J., … Tengström, A. (2015). The Effectiveness of Acceptance and Commitment Therapy for Adolescent Mental Health: Swedish and Australian Pilot Outcomes. *Journal of Child and Family Studies*, *24*, 1016–1030. http://doi.org/10.1007/s10826-014-9912-9

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *3*(3), 335–343.

MacKay, T. (2007). Educational psychology: The fall and rise of therapy. *Educational and Child Psychology*, *24*(1), 7–18.

Maxwell, S. E. (2004). The Persistence of Underpowered Studies in Psychological Research : Causes , Consequences , and Remedies, *9*(2), 147–163. http://doi.org/10.1037/1082-989X.9.2.147

Morris, S. B. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, *11*(2), 364–386. http://doi.org/10.1177/1094428106291059

Murphey, D., Barry, M., & Vaughn, B. (2013). Mental Health Disorders. *Child Trends*, (January), 1–10. Retrieved from http://www.cdc.gov/nchs/fastats/mental.htm

Murrell, A. R., Steinberg, D. S., Connally, M. L., Hulsey, T., & Hogan, E. (2015). Acting Out to ACTing On: A Preliminary Investigation in Youth with ADHD and Co-morbid Disorders. *Journal of Child and Family Studies*, *24*(7), 2174–2181. http://doi.org/10.1007/s10826-014-0020-7

Murrell, A. R., & Wilson, K. G. (2002). *ACT with Children and Adolescents Manual*.

Pahnke, J., Lundgren, T., Hursti, T., & Hirvikoski, T. (2014). Outcomes of an acceptance and commitment therapy-based skills training group for students with high-functioning autism spectrum disorder: A quasi-experimental pilot study. *Autism : The International Journal of Research and Practice*, *18*(8), 953–964. http://doi.org/10.1177/1362361313501091

Selye, H. (1974). *Stress without distress*. Philadelphia: J.B. Lippincott Company.

Swain, J., Hancock, K., Dixon, A., & Bowman, J. (2015). Acceptance and Commitment Therapy for children: A systematic review of intervention studies. *Journal of Contextual Behavioral Science*, *4*(2), 73–85. http://doi.org/10.1016/j.jcbs.2015.02.001

Szabó, M., & Lovibond, P. F. (2006). Anxiety, depression, and tension/stress in children. *Journal of Psychopathology and Behavioral Assessment*, *28*(3), 195–205. http://doi.org/10.1007/s10862-005-9008-3

Theodore-Oklota, C., Orsillo, S. M., Lee, J. K., & Vernig, P. M. (2014). A pilot of an acceptance-based risk reduction program for relational aggression for adolescents. *Journal of Contextual Behavioral Science*, *3*(2), 109–116. http://doi.org/10.1016/j.jcbs.2014.03.001

Wade, C. (2016). Therapeutic practice within educational psychology: The discursive contribution of therapeutic practice from the perspective of educational psychologists new to the profession. *Educational and Child Psychology*, *33*(4), 8–24.

Waters, C. S. (2012). *Acceptance and Commitment Therapy: Efficacy and Mechanisms of Therapeutic Action*. University of Cardiff. Retrieved from http://core.kmi.open.ac.uk/download/pdf/9713295.pdf%5Cnhttp://orca.cf.ac.uk/3

7760/1/DCLIN PSYCH Thesis Dr Cerith Waters.pdf

Wheaton, B., & Montazer, S. (2012). Stressors , Stress , and Distress. In T. L. Scheid & T. N. Brown (Eds.), *A Handbook for the Study of Mental Health Social Contexts, Theories, and Systems* (2nd ed., pp. 171–199). Cambridge: Cambridge University Press.

Wolpert, M., Deighton, J., Patalay, P., Martin, A., Fitzgerald-Yau, N., Demir, E., … Meadows, P. (2011). *Me and My School: Findings from the National Evaluation of Targeted Mental Health in*. London. Retrieved from http://discovery.ucl.ac.uk/1391832/

Wolpert, M., Fugard, A., & Deighton, J. (2013). Issues in evaluation of psychotherapies. In P. Graham & S. Reynolds (Eds.), *Cognitive Behaviour Therapy for Children and Families* (3rd ed., pp. 34–47). Cambridge: Cambridge University Press.

Appendix A

Table 8
Excluded Articles

Reference	Reason for Exclusion
Armstrong, A. B., Morrison, K. L., & Twohig, M. P. (2013). A Preliminary Investigation of Acceptance and Commitment Therapy for Adolescent Obsessive-Compulsive Disorder. *Journal of Cognitive Psychotherapy*, *27*(2), 175–190. http://doi.org/10.1891/0889-8391.27.2.175	2. Study not set in a school
Azadeh, S. M., Kazemi-Zahrani, H., & Besharat, M. A. (2015). Effectiveness of Acceptance and Commitment Therapy on Interpersonal Problems and Psychological Flexibility in Female High School Students With Social Anxiety Disorder. *Global Journal of Health Science*, *8*(3), 131–138. http://doi.org/10.5539/gjhs.v8n3p131	2. Study not set in a school
Brown, F. J., & Hooper, S. (2009). Acceptance and Commitment Therapy (ACT) with a learning disabled young person experiencing anxious and obsessive thoughts. *Journal of Intellectual Disabilities*, *13*(3), 195–201. http://doi.org/10.1177/1744629509346173	2. Study not set in a school
Cattivelli, R., Cavallini, F., Tirelli, V., Maffini, N., & Perini, S. (2012). Educational outlook through a clinical approach: The contributions of Acceptance and Commitment Therapy and of Functional Analytic Psychotherapy in the case of a boy with social anxiety. *Psicoterapia Cognitiva e Comportamentale. 18(3), 357-368.*	7. Language of report is not English
Dixon, M. R. (2013). Don't stop believing: journeys school. *Behavior Analysis in Practice*, *6*(1), 78–79.	5. Study does not have outcome variable measuring psychological distress
Gómez, M. J., Luciano, C., Páez-Blarrina, M., Ruiz, F. J., Valdivia-Salas, S., & Gil-Luciano, B. (2014). Brief ACT protocol in at-risk adolescents with conduct disorder and impulsivity. *International Journal of Psychology and Psychological Therapy*, *14*(3), 307–332.	5. Study does not have outcome variable measuring psychological distress

Reference	Reason for Exclusion
Hancock, K. M., Swain, J., Hainsworth, C. J., Dixon, A. L., Koo, S., & Munro, K. (2016). Acceptance and commitment therapy versus cognitive behavior therapy for children with anxiety: outcomes of a randomized controlled trial. *Journal of Clinical Child and Adolescent Psychology*, *4416*(August), 1–16. http://doi.org/10.1080/15374416.2015.1110822	2. Study not set in a school
Luciano, C., Ruiz, F. J., & Vizcaíno-Torres, R. (2011). A relational frame analysis of Defusion interactions in Acceptance and Commitment Therapy. A preliminary and quasi-experimental study with at-risk …. International Journal of …, 165–182.	5. Study does not have outcome variable measuring psychological distress
Ruiz, F. J., & Luciano, C. (2009). Eficacia de la terapia de aceptación y compromiso (ACT) en la mejora del rendimiento ajedrecístico de jóvenes promesas. [Acceptance and commitment therapy (ACT) and improving chess performance in promising young chess-players.]. *Psicothema*, *21*(3), 347–352.	7. Language of report is not English
Woidneck, M. R., Morrison, K. L., & Twohig, M. P. (2014). Acceptance and commitment therapy for the treatment of posttraumatic stress among adolescents. *Behavior Therapy*, *38*(4), 451–476. http://doi.org/10.1177/0145445513510527	2. Study not set in a school

Appendix B

Table 9
Summary of Included Studies

Authors	N	Participants	Design	Measures of Psychological Distress	Intervention	Comparison condition	Adults delivering intervention
Burckhardt (2016)	63	Australian independent high school students aged 15-18. High SES.	RCT. Cluster randomised by tutorial group.	Depression, Anxiety and StressScale-short form (DASS-21).	'Strong minds'-school-based mental health program combining positive psychology with ACT. 8hrs programme, comprising 16 half hour sessions over 3 months. Mostly workshops were twice per week. Delivered in large groups in amphitheatre.	Pastoral care class as usual. Teaching about social justice, cyber issues, drugs, safe behaviour. Length, duration and total number of sessions same as intervention group.	Clinical psychologist (lead author) supervised by experienced ACT practitioners.
Livheim (2015) A (Australian)	66	Australian high school students aged 12-18. Mostly female students (only 8 male students).	RCT. Individual randomisation of female students to group but no randomisation of male students who were all assigned to ACT intervention. .	Reynolds Adolescent Depression Scale-2 (RADS-2)	ACT Experiential Adolescent Group-manualized 8-week group program (Hayes & Rowse, 2008). Uses experiential mediums (painting and role-play) to help adolescents understand ACT processes.	12 weeks of monitoring support by school counsellor (standard care for students identified as at risk).	Registered psychologists and co-facilitated by clinical psychology graduate students or school's own counsellor. Staff received minimum 2 days ACT training and supervised by authors of programme (Hayes and Rowse).

Authors	N	Participants	Design	Measures of Psychological Distress	Intervention	Comparison condition	Adults delivering intervention
Livheim (2015) B Swedish	32	Swedish high school students aged 14-15. Mostly female students (only 9 male students). Needed above 80% on scales measuring psychological problems but excluded if had severe psychiatric problems. High SES	RCT. Randomisation at individual level- stratified by gender to create 1 male group.	Perceived Stress Scale (PSS). Translated into Swedish. Depression, Anxiety and StressScale-short form (DASS-21). Translated into Swedish.	ACT Experiential Adolescent Group-manualized 8-week group program (Hayes & Rowse, 2008). Uses experiential mediums (painting and role-play) to help adolescents understand ACT processes.Translated into Swedish. 8 group sessions fitted into 6-week period.	Individual support by school nurse. Half of students in control group met with nurse for individual counselling (between 2-8 sessions).	Two clinical psychology major students with clinical training in CBT and 4 days ACT training.
Murrell (2015)		9 children aged 11-15. 5 females, 4 males at urban charter school in south central US. All study participants met criteria for ADHD and some also had comorbid LDs or behaviour problems.	Small N design	Emotional Symptoms Index of BASC-2 (Behaviour Assessment System for Children, second edition). This index is composed of scores on the Social Stress, Anxiety and Depression scales.	1hr group sessions based on ACT for kids protocol developed by Murrell and Wilson, (2002)Once per week for 8 weeks. Descriptions of group sessions included in paper. Used ACT techniques to reduce experiential avoidance and increase valuing. Metaphors, 'concrete enactments' and art projects used. Discussions about thoughts, behaviours and feelings.	n/a	Recognized ACT trainer with assistance from graduate and advanced undergraduate students.

Authors	N	Participants	Design	Measures of Psychological Distress	Intervention	Comparison condition	Adults delivering intervention
Theodore-Oklota 2014		210 seventh grade students with mean age of 12.45 years (range of ages not given) at a suburban middle school in New England, USA.	One teaching cohort of four classrooms was randomised to receive programme and other teaching cohort of four classrooms was waitlisted to receive programme after 3 months.	Child behavior checklist – Youth self-report- checklist of behavioural and emotional functioning. Some items were removed due to use of passive consent. Total problems score used as indicator of overall psychological distress.	3 sessions occurring over course of 2 weeks during a 48-minute class period. Involved group discussions, role-playing and experiential exercises. Session 1- discussion of relational aggression and values Session 2- coping strategies e.g. avoidant coping, passengers on bus analogy. Session 3- Role-play on relationships, looking at values and coping. Following completion of programme, packet of information with further reading and exercises given.	Waitlist control group.	Doctoral students trained in delivery of the programme (during 2-3 six-hour sessions). Met with first author after each session to troubleshoot.

Authors	N	Participants	Design	Measures of Psychological Distress	Intervention	Comparison condition	Adults delivering intervention
Pahnke (2016)	28	13-21 year olds in a special school. All with high functioning ASD. Stockholm, Sweden.	Cluster randomised intervention/waiting list control. Repeated measures. Classes were randomised to a condition. No individual randomisation.	Stress Survey Scale (teacher and self-ratings) Strengths and Difficulties Questionnaire (teacher and self-ratings) Beck Youth Inventories (self-ratings)	6 week ACT based skills training programme. Two 40 min group sessions and 6-12 minutes of mindfulness exercises in the classroom per day. Small groups.	Classes as usual (waiting list control).	Mindfulness training sessions were facilitated by class teacher. Group sessions were led by graduate student (1st author) supervised by ACT therapist (2nd author)

Appendix C
Weight of Evidence Ratings

Weight of Evidence A

Group Studies

Weight of Evidence A was determined by using adapted non-review specific protocols based on the work of Kratochwill et al., (2003) for group studies and Horner et al. (2005) for the small N case design.

The Kratochwill et al., (2003) protocol was adapted to suit this review by removing sections on identifiable components, replication, site of implementation and follow-up assessment.

Table 9
Alterations to Kratochwill et al., (2003) Coding Protocol

Section removed from protocol	Rationale for removal
Identifiable components	No studies were able to identify which intervention components were effective. Kratochwill et al. (2003) include this criterion to influence direction of future research.
Replication	Review will consider all studies and whether their findings coincide throughout.
Site of implementation	Inclusion criteria require that studies are all located in school.
Follow-up assessment	Only one study included a follow-up assessment.
Outcomes (statistical analysis part of this section retained)	Outcomes are addressed separately in effect size table.

Table 10
Weight of Evidence A Criteria for group studies

Criteria	Ratings
A: Measurement	3. Studies use measures with reliability of .85 or higher. Data is collected from multiple sources and using multiple methods. A case for validity must be presented.
	2. Measures have reliability of .70 or higher. Data is collected using either multiple sources or multiple methods.
	1. Measures have reliability of .50 or higher.
B: Comparison group	3. Active comparison group used. These must all be present: counterbalancing of change agents, group equivalence established, equivalent mortality with low attrition.
	2. Comparison group used but may not be 'active' e.g. waitlist. Two of these three must be present: counterbalancing of change agents, group equivalence established, equivalent mortality with low attrition.
	1. Comparison group used and one of the following present: counterbalancing of change agents, group equivalence established, equivalent mortality with low attrition.
C: Statistical Analysis	3. Study must use appropriate units of analysis, familywise/experimentwise error rate controlled (if applicable), and a sufficiently large N. 75% of primary outcomes must be significant.
	2. Study must use appropriate units of analysis, familywise/experimentwise error rate controlled (if applicable), and a sufficiently large N.
	1. Study must use appropriate units of analysis and familywise/experimentwise error rate controlled (if applicable).

Criteria	Ratings
D: Educational/clinical significance	3. Study must provide evidence in support of the clinical significance for at least 3 of the 4 criteria listed (i.e., Categorical Diagnosis Data, Outcomes Assessed Via Continuous Variables, Subjective Evaluation, or Social Comparison) during either post or follow up phases for the majority of participants 2. Study must provide evidence in support of the clinical significance for at least 2 of the 4 criteria listed (i.e., Categorical Diagnosis Data, Outcomes Assessed Via Continuous Variables, Subjective Evaluation, or Social Comparison) during either post or follow up phases for the majority of participants 1. Study must provide evidence in support of the clinical significance for at least 1 of the 4 criteria listed (i.e., Categorical Diagnosis Data, Outcomes Assessed Via Continuous Variables, Subjective Evaluation, or Social Comparison) during either post or follow up phases for the majority of participants.
E: Fidelity	3. Evidence of adherence should be measured through at least two of the following: ongoing supervision/consultation, coding sessions, or audio/video tapes, and use of a manual. Information must have been provided to the implementers using either: (1) written materials involving a detailed account of the exact procedures and the sequence in which they are to be used or (2) a formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used. There is a description of any adaptations. 2. Evidence of adherence should be measured through at least one of the above criteria and use of a manual. To be considered a “manual” for a rating of 2, information must have been provided to the implementers using either: (1) written materials involving an overview of broad principles and a description of the intervention phases, or (2) a formal or informal training session involving an overview of broad principles and a description of the intervention phases 1. Evidence of acceptable adherence should be measured through at least one of the above criteria or use of a manual.

Table 11
Weight of Evidence A Ratings (Group Studies)

Authors	Dimensions					Total
	A	B	C	D	E	WoE A
Burckhardt et al. 2016	1	2	1	1	3	1.6
Livheim et al. 2015a	1	2	1	0	2	1.2
Livheim et al. 2015b	2	2	1	0	1	1.2
Pahnke et al. 2014	0	2	1	0	1	0.8
Theodore-Oklota et al. 2014	1	2	2	0	1	1.2

Small N study

The Horner et al. (2005) protocol was used in full. This was used for the Murrell et al. (2015) study as it employed a small-N case design. The quality indicators detailed in Table 1 of the Horner et al. (2005, p175) paper were used as criteria for Weight of Evidence A. The descriptors used by Horner and colleagues were used to give different possible numerical ratings. To arrive at a total Weight of Evidence A for the small N study, ratings for all criteria were summed and divided by 7, which was the total number of criteria.

Table 12
Weight of Evidence A Criteria for small-N study

Criteria	Ratings
A: Description of Participants and Setting	3. All 3 of the following criteria are fulfilled: participants are described with sufficient detail to allow others to select individuals with similar characteristics (e.g., age, gender, disability, diagnosis); the process for selecting participants is described with replicable precision; critical features of the physical setting are described with sufficient precision to allow replication. 2. Two out of the criteria above are fulfilled. 1. One of the criteria above is fulfilled.
B: Dependent Variable	3. All of the following criteria are fulfilled: dependent variables are described with operational precision; each dependent variable is measured with a procedure that generates a quantifiable index; measurement of the dependent variable is valid and described with replicable precision; dependent variables are measured repeatedly over time; data are collected on the reliability or interobserver agreement associated with each dependent variable, and IOA levels meet minimal standards (e.g., IOA = 80%; Kappa = 60%). 2. Three to four of the above criteria are fufilled. 1. One to two of the above criteria are fulfilled.
C: Independent Variable	3. All of the following criteria are fulfilled: independent variable is described with replicable precision; independent variable is systematically manipulated and under the control of the experimenter; overt measurement of the fidelity of implementation for the independent variable is highly desirable. 2. Two of the above criteria are fulfilled. 1. One of the above criteria is fulfilled.

D: Baseline	3. All of the following criteria are fulfilled: the study includes a baseline phase that provides repeated measurement of a dependent variable; the study establishes a pattern of responding that can be used to predict the pattern of future performance, if introduction or manipulation of the independent variable did not occur; baseline conditions are described with replicable precision. 2. Two of the above criteria are fulfilled. 1. One of the above criteria is fulfilled.
E: Experimental Control/Internal Validity	3. All of the following criteria are fulfilled: the design provides at least three demonstrations of experimental effect at three different points in time; the design controls for common threats to internal validity (e.g., permits elimination of rival hypotheses); the results document a pattern that demonstrates experimental control. 2. Two of the above criteria are fulfilled. 1. One of the above criteria is fulfilled.
F: External Validity	3. Experimental effects are replicated across 3 or more participants and also either settings or materials to establish external validity. 2. Experimental effects are replicated across 3 or more participants 1. Experimental effects are replicated by inclusion of at least 2 participants.
G: Social validity	3. All of the following criteria are fulfilled: the dependent variable is socially important; the magnitude of change in the dependent variable resulting from the intervention is socially important; implementation of the independent variable is practical and cost effective; social validity is enhanced by implementation of the independent variable over extended time periods, by typical intervention agents, in typical physical and social contexts. 2. Two to three of the above criteria are fulfilled. 1. One of the above criteria is fulfilled.

Table 13
Weight of Evidence A Ratings (Small N Study)

Authors	Dimensions							Total WoE A
	A	B	C	D	E	F	G	
Murrell et al. 2015.	1	2	0	0	0	1	1	0.9

Weight of Evidence B

Weight of Evidence B was determined by review-specific criteria relating to methodological relevance. Different criteria were used for group studies and single case designs. To receive weightings of two or three, the study must have employed a group design.

Table 14
Weight of Evidence B Criteria

Criteria	Ratings	Rationale
A: Randomisation procedure	3. Group study with full individual randomisation of participants to intervention or control group	Full individual randomisation allows conclusions about causality to be drawn more convincingly. Pre-existing groups may differ systematically.
	2. Group study with cluster randomisation (e.g. by classroom) or partial individual randomisation.	
	1. Group study with no randomisation.	
B: Comparison group	3. Comparison group also receives psychotherapeutic support in the same format as intervention group e.g. CBT, other talk therapy	If other psychotherapeutic support is offered to control group in same format, more secure conclusions can be drawn that it is the ACT component of the intervention that is bringing about change. If not, positive effects of programme may be due to social support in

	2. Comparison group receives an intervention in same format as intervention group but this may not be psychotherapy. 1. Comparison group receives little or no alternative intervention.	group or attention.

Ratings for criteria A and B were summed and divided by two to arrive at a score for Weight of Evidence B.

Table 15
Weight of Evidence B Ratings (Group Studies)

Authors	Dimensions		Total WoE B
	A	B	
Burckhardt et al. 2016	3	2	2.5
Livheim et al. 2015a	2	1	1.5
Livheim et al. 2015b	3	1	2
Pahnke et al. 2014	2	1	1.5
Theodore-Oklota et al. 2014	2	1	1.5

The Murrell et al., (2015) study was a small N design and was assigned an overall WoE B of one since it did not employ randomisation or a comparison group.

Weight of Evidence C

Weight of Evidence C was determined by review-specific criteria relating to topic relevance.

Table 16
Weight of Evidence C Criteria

Criteria	Ratings	Rationale
A: ACT components	3. All components of ACT are evident in the intervention	Review attempts to test the effectiveness of ACT-based

	(cognitive defusion, acceptance, contact with the present moment, observing the self, values and committed action). 2. At least 4 components of ACT are evident in the intervention 1. 1-3 components of ACT are evident in the intervention or information on intervention components is not stated	interventions, therefore interventions with more components of ACT will provide better evidence.

Criteria	Ratings	Rationale
B: Presenting problems of sample	3. Sample is screened and found to be experiencing mild-moderate psychological distress (depression, anxiety or stress) according to diagnostic criteria or a valid and reliable scale 2. Sample is nominated by staff as experiencing or likely to be experiencing mild-moderate psychological distress (depression, anxiety or stress) but this has not been validated by a diagnostic process, psychological assessment or screening tool. OR Sample has a diagnosed psychological condition e.g. ASD and authors make a case for children and young people with this condition commonly experiencing psychological distress 1. Sample has no identified psychological condition or authors do not link the psychological condition they have to psychological distress	
C: Intention of Intervention	3. Intervention was planned to reduce psychological distress (stress, anxiety, depression)	

	2. Intervention was planned to help a different psychological condition 1. No clear rationale for intervention	

Criteria	Ratings	Rationale
D: Generalisability of sample and setting	3. Sample demographics (sex and socioeconomic status) of participants is described in detail and both characteristics are representative of the general population in that country. AND Setting is typical of that attended by the general population in that country. 2. Sample demographics (sex and socioeconomic status) are described in detail and one of these characteristics is representative of the general population in that country. AND Setting is typical of that attended by the general population in that country. 1. Sample demographics are not typical for that country or not well described. OR Setting is not typical of that attended by the general population in that country (independent school or special school).	

Table 17
Weight of Evidence C Ratings

Authors	Dimensions				Total WoE C
	A	B	C	D	
Burckhardt et al. 2016	3	3	3	1	2.5
Livheim et al. 2015a	3	2	3	1	2
Livheim et al. 2015b	3	3	3	1	2.5
Murrell et al., 2015	3	1	2	1	1.8
Pahnke et al. 2014	3	2	3	1	2
Theodore-Oklota et al. 2014	1	1	2	2	1.5

Ratings for criteria A, B and C were summed and divided by three to arrive at a score for Weight of Evidence C.

Weights of Evidence A, B and C were summed and divided by 3 to arrive at a total score for Weight of Evidence D.

Appendix D

Example of Coding Protocol for Group Designs (Kratochwill et al., 2003)

1. Burckhardt et al. (2016)

Date 05.02.2017

Full Study Reference in proper format:

(Burckhardt et al., 2016)

Burckhardt, R., Manicavasagar, V., Batterham, P. J., & Hadzi-Pavlovic, D. (2016). A randomized controlled trial of strong minds: A school-based mental health program combining acceptance and commitment therapy and positive psychology. *Journal of School Psychology*, *57*, 41–52. http://doi.org/10.1016/j.jsp.2016.05.008

☐ Type of Publication:
☐ Book/Monograph
☒ Journal Article
☐ Book Chapter
☐ Other (specify):

1. General Characteristics

A. General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)

☐ Completely randomised design
☐ Randomised block design (between participants, e.g., matched classrooms)
☐ Randomised block design (within participants)
☐ Randomised hierarchical design (nested treatments
☒ Cluster randomisation (8 tutorial groups randomised to intervention group, 8 to control group)

A2. Nonrandomised designs (if non-random assignment design, select one of the following)

☐ Nonrandomised design
☐ Nonrandomised block design (between participants)
☐ Nonrandomised block design (within participants)
☐ Nonrandomised hierarchical design
☐ Optional coding for quasi-experimental designs

A3. Overall confidence of judgment on how participants were assigned (select one of the following)

☐ Very low (little basis)
☐ Low (guess)
☐ Moderate (weak inference)
☐ High (strong inference)
☒ Very high (explicitly stated)
☐ N/A
☐ Unknown/unable to code

B. Participants

Total size of sample (start of study): __63_

Intervention group sample size:___30_

Control group sample size:___33___

C. Type of Program

☒ Universal prevention program
☐ Selective prevention program
☐ Targeted prevention program
☐ Intervention/Treatment
☐ Unknown

D. Stage of Program

☒ Model/demonstration programs
☐ Early stage programs
☐ Established/institutionalized programs
☐ Unknown

E. Concurrent or Historical Intervention Exposure

☐ Current exposure
☐ Prior exposure
☒ Unknown

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

(Rating Scale: **3= Strong Evidence, 2=Promising Evidence, 1=Weak Evidence, 0=No Evidence)**

A. Measurement (Estimating the quality of the measures used to establish effects)

A1 The use of the outcome measures produce reliable scores for the majority of the primary outcomes

☒Yes
☐No
☐Unknown/unable to code

A2 Multi-method (at least two assessment methods used)
☐ Yes
☒ No (different scales were used but to assess different constructs e.g. depression, anxiety, stress, "flourishing")
☐ N/A
☐ Unknown/unable to code

A3 Multi-source (at least two sources used self-reports, teachers etc.)
☐ Yes
☒ No
☐ N/A
☐ Unknown/unable to code

A4 Validity of measures reported (well-known or standardized or norm-referenced are considered good, consider any cultural considerations)

☐ Yes validated with specific target group
☒ In part, validated for general population only
☐ No

☐ Unknown/unable to code

Overall Rating for measurement_____1_____

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0=No Evidence

B. Comparison Group

B1 Type of Comparison Group (Select one of the following)

☒ Typical intervention (typical intervention for that setting, without additions that make up the intervention being evaluated) Usual pastoral care classes
☐ Attention placebo
☐ Intervention element placebo
☐ Alternative intervention
☐ Pharmacotherapy
☐ No intervention
☐ Wait list/delayed intervention
☐ Minimal contact
☐ Unable to identify type of comparison

B2 Overall confidence of judgment on type of comparison group

☐ Very low (little basis)
☐ Low (guess)
☐ Moderate (weak inference)
☐ High (strong inference)
☒ Very high (explicitly stated)
☐ Unable to identify comparison group

B3 Counterbalancing of change agent (participants who receive intervention from a single therapist/teacher etc. were counter-balanced across intervention)

☐ By change agent
☐ Statistical (analyse includes a test for intervention)
☐ Other
☒ Not reported/None School staff ran pastoral care classes

B4 Group equivalence established (select one of the following)

☐ Random assignment
☐ Posthoc matched set
☐ Statistical matching
☒ Post hoc test for group equivalence- no significant differences on baseline scores or demographics found

B5 Equivalent mortality
☒ Low attrition (less than 20 % for post) (9% in control condition and 8.2% in intervention condition)
☐ Low attrition (less than 30% for follow-up)
☐ Intent to intervene analysis carried out?
Findings_____

Overall rating for Comparison group _____2_____

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0=No Evidence

C. Appropriate Statistical Analysis

Analysis 1_____DASS_____

☒ Appropriate unit of analysis
☒ Familywise/experimenter wise error rate controlled when applicable
☐ Sufficiently large N- (large sample used but only 63 for DASS because they only looked at the students who began programme with elevated symptoms)

Overall rating for statistical analysis _____1_____

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0=No Evidence

D. Educational/Clinical Signficance

Study must provide evidence in support of the clinical significance for:

☐ Categorical Diagnosis Data
☒ Outcomes Assessed Via Continuous Variables
☐ Subjective Evaluation
☐ Social Comparison

Overall rating for educational/clinical significance _____1_____

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0=No Evidence

E. Fidelity

Evidence of adherence is measured through

☒ Ongoing supervision/consultation
☒ Coding sessions
☒ Audio/video tapes
☐ Use of a manual
☐ There is a description of any adaptations.

Overall rating for fidelity _____3_____

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0=No Evidence

Summary of Evidence

Indicator	**Overall evidence rating 0-3**	**Description of evidence Strong Promising Weak No/limited evidence Or Descriptive ratings**
General Characteristics		
Design		Cluster Randomisation
Type of programme		Universal prevention programme
Stage of programme		Model/demonstration programme
Concurrent/ historical intervention exposure		Unknown
Key features		
Measurement	1	Weak evidence
Comparison group	2	Promising evidence
Appropriate Statistical Analysis	1	Weak evidence
Educational/Clinical Significance	1	Weak Evidence
Fidelity	3	Strong Evidence

Coding protocol for Single Case Research (Horner et al., 2005.)

See Table 12 for explanation of criteria used to rate this study.

Murrell et al. (2015)

Murrell, A. R., Steinberg, D. S., Connally, M. L., Hulsey, T., & Hogan, E. (2015). Acting Out to ACTing On: A Preliminary Investigation in Youth with ADHD and Co-morbid Disorders. *Journal of Child and Family Studies*, *24*(7), 2174–2181. http://doi.org/10.1007/s10826-014-0020-7

Section A: Description of Participants and Setting

1) Participants are described with sufficient detail to allow others to select individuals with similar characteristics; (e.g., age, gender, disability, diagnosis).

Yes
No – they all have ADHD but also 'comorbid LDs and/or disruptive behaviour'- not specified exactly what. Gender and age are given.
N/A
Unknown/Unable to Code

2) The process for selecting participants is described with operational precision.

Yes
No – they 'volunteered'
N/A
Unknown/Unable to Code

3) Critical features of the physical setting are described with sufficient precision to allow replication.

Yes
No
N/A
Unknown/Unable to Code

Overall Rating of Evidence: 1

Section B: Dependent Variable

1) Dependent variables are described with operational precision.

Yes- Behavior Assessment System for Children, 2 (BASC-2).
No
N/A
Unknown/Unable to Code

2) Each dependent variable is measured with a procedure that generates a quantifiable index.

Yes
No
Unknown/Unable to Code

3) Measurement of the dependent variable is valid and described with replicable precision.

Yes
No
N/A
Unknown/Unable to Code

4) Dependent variables are measured repeatedly over time.

Yes
No
N/A
Unknown/Unable to Code

5) Data are collected on the reliability or inter-observer agreement associated with each dependent variable, and IOA levels meet minimal standards (e.g., IOA = 80%; Kappa = 60%).

Yes – see BASC 2 article
No
N/A
Unknown/Unable to Code

Overall Rating of Evidence: 2

Section C: Independent Variable

1) Independent variable is described with replicable precision.

Yes
No- details of intervention are given but not enough to replicate the study precisely
N/A
Unknown/Unable to Code

2) Independent variable is systematically manipulated and under the control of the experimenter.

Yes
No
N/A
Unknown/Unable to Code

3) Overt measurement of the fidelity of implementation for the independent variable is highly desirable.

Yes
No
N/A
Unknown/Unable to Code

Overall Rating of Evidence: 0

Section D: Baseline

1) The majority of single-subject research studies will include a baseline phase that provides repealed measurement of a dependent variable and establishes a pattern of responding that can be used to predict the pattern of future performance, if introduction or manipulation of the independent variable did not occur.

Yes
No

N/A
Unknown/Unable to Code

2) Baseline conditions are described with replicable precision.

Yes
No
N/A
Unknown/Unable to Code

Overall Rating of Evidence: 0

Section E: Experimental Control/Internal Validity

1) The design provides at least three demonstrations of experimental effect at three different points in time.

Yes
No
N/A
Unknown/Unable to Code

2) The design controls for common threats to internal validity (e.g., permits elimination of rival hypotheses).

Yes
No
N/A
Unknown/Unable to Code

3) The results document a pattern that demonstrates experimental control.

Yes
No – no effect was found
N/A
Unknown/Unable to Code

Overall Rating of Evidence: 0

Section F: External Validity

1) Experimental effects are replicated across participants, settings, or materials to establish external validity.

Yes - multiple participants were used but no effect found

No
N/A
Unknown/Unable to Code

Overall Rating of Evidence: 2

Section G: Social Validity

1) The dependent variable is socially important.
Yes- emotional distress
No
N/A
Unknown/Unable to Code

2) The magnitude of change in the dependent variable resulting from the intervention is socially important.
Yes
No- no change effect found
N/A
Unknown/Unable to Code

3) Implementation of the independent variable is practical and cost effective
Yes
No- delivered by researcher trained in ACT
N/A
Unknown/Unable to Code

4) Social validity is enhanced by implementation of the independent variable over extended time periods, by typical intervention agents, in typical physical and social contexts.
Yes
No- researcher trained in ACT needed, lots of rewards necessary to get students to go to group, teachers did not fill in Qaires
N/A
Unknown/Unable to Code

Overall Rating of Evidence: 1

A: 1
B: 2
C: 0
D: 0
E: 0
F: 2
G: 1

Average WoE A = 0.9
Sum of X / N
X = individual quality rating for each judgement area
N = number of judgement areas