

VOLUME ONE

EXPLORING THE EFFECTIVENESS OF ACCEPTANCE AND COMMITMENT
THERAPY AND ITS UNDERLYING PROCESSES

By

LOIS LINDA CARTER

**A THESIS SUBMITTED TO THE UNIVERSITY OF BIRMINGHAM FOR THE
DEGREE OF DOCTOR OF CLINICAL PSYCHOLOGY**

Department of Clinical Psychology

School of Psychology

The University of Birmingham

25th May 2018

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Thesis Overview

The thesis consists of two volumes and is submitted in partial fulfilment of the Doctorate of Clinical Psychology. Volume One is the research component consisting of a systematic literature review, and empirical study paper and a public dissemination document. Volume Two is the clinical component consisting of five clinical practice reports.

VOLUME ONE

The first paper in Volume One is a meta-analysis investigating the effectiveness of Acceptance and Commitment Therapy interventions with children and adolescents across a range of conditions. The second paper is an empirical study exploring the use of the Implicit Relational Assessment Procedure as a measure of cognitive fusion and self-as-context processes used in ACT. The third paper provides an executive summary for the dissemination of findings to the public and relevant stakeholders.

VOLUME TWO

The first report in Volume Two is case formulation with a 67-year-old lady with depression from a Cognitive-Behavioural and Psychodynamic perspective. The second report is a service evaluation investigating access rates for older adults in an Improving Access to Psychological Therapy (IAPT) service and the potential barriers of access for this group. The third paper is a case study with a 52-year-old lady with depression and interpersonal difficulties using a psychodynamic informed approach. The fourth paper presents a single-case experimental in a Cognitive-Behavioural Therapy intervention with a 12-year-old girl with anger outbursts. The final report presents a short-term psychodynamic approach with a 54-year-old lady with chronic pain and depression.

Acknowledgments

I would like to thank the Clinical Psychology Team at the University of Birmingham for their continued support and guidance. I would like to thank Dr Richard Bennett who supervised the research project in Volume One of the thesis whose enthusiasm for Acceptance and Commitment Therapy was infectious. I would like to thank Dr Chris Jones who supervised my meta-analysis. He has offered great dedication and support, pushing me to reach my best until the very end. Chris' unswerving passion for meta-analyses has also rubbed off on me a little and I have learnt a number of word processing skills along the way. I would also like to thank Dr Andy Fox who co-supervised the meta-analysis who was always willing to help and do what he could to support. Finally, I would like to thank the calming presence of my appraisal tutor, Dr Gerry Riley, who has always been there and on-hand for any last-minute panics before deadlines.

I would like to thank my fellow trainees who it has been a pleasure to get to know over the past three years and has made the journey a lot more enjoyable. I would like to thank all my patients, participants and health professionals without whom this would not have all been possible. I would also like to thank all of my supervisors; Dr Mike Ridley-Dash, Dr Rachel Spector, Dr Brigid Duffy, Dr Adam Cleary and Dr Luke Brown. I will take a little bit of each of you forward in my development as a Clinical Psychologist.

I would also like to thank Ann for her exceptional eye for detail and her support in this process. I would also like to thank my parents, family and friends who have supported me along the way who have helped me to endure the stressful times, offering their kind words and helpful distractions when needed. Lastly, I would like to thank Dom, my Greyhound for keeping me company during the study block.

VOLUME ONE

List of Contents

CHAPTER I: The Effectiveness of Acceptance and Commitment Therapy with Children and Adolescents: A Meta-Analysis

1. Abstract	2
1.1 Background	2
1.2 Method	2
1.3 Results	2
1.4 Conclusion.....	3
2. Introduction	4
2.1 Overview of ACT.....	4
2.3 Rationale for Investigating the Effectiveness of ACT with Children and Adolescents	5
2.4 Previous Reviews looking at ACT for Children and Adolescents	6
2.5 Moderators of therapeutic efficacy.....	7
2.6 Limitations of Previous Reviews	7
2.7 Aims	8
3. Method	9
3.1 Search Strategy.....	9
3.2 Inclusion/exclusion criteria	10
3.3 Quality Criteria.....	17
3.3.1 Methodological Bias	20
3.3.2 Measurement of Outcomes.....	20
3.3.3 Blinding of Outcome Assessors	20
3.3.4 Selective Reporting	21
3.3.5 Treatment Fidelity	21
3.3.6 Treatment Validity	21
3.3.7 Attrition Bias	22
3.3.8 Power.....	22
3.3.9 Summary	22

3.4	Data Extraction.....	23
3.5	Data Analysis	26
3.6	Meta-analytic model.....	26
3.7	Quality Effects.....	27
3.8	Quantifying Heterogeneity.....	27
3.9	Identifying Influential Studies.....	27
3.9.1	One Left Out Procedure	27
3.9.2	Influential Case Diagnostics.....	28
3.10	Publication Bias.....	29
3.11	Moderator Analyses	29
4.	Results	30
4.1	Participant Characteristics.....	30
4.2	Intervention Characteristics.....	30
4.3	Summary	31
4.4	Depression.....	33
4.4.1	Post Intervention	33
4.4.3	Follow-up	34
4.5	Anxiety.....	36
4.5.1	Post Intervention	36
4.5.2	Follow-up	37
4.6	Psychological Flexibility.....	39
4.7	Functioning	40
4.7.1	Self-reports.....	40
4.7.2	Follow-up	41
4.7.3	Reported by Parents	42
4.8	Potential Moderating Variables.....	43
5.	Discussion	50
5.1	Summary of findings.....	50
5.2	Potential Moderating Variables.....	52
5.3	Clinical Implications	54
5.4	Convergence with Other Reviews.....	56
5.5	Recommendations for Future Research	57
5.6	Conclusion.....	57
	References	59

CHAPTER II: The Empirical research Report: Exploring the Implicit Relational Assessment Procedure (IRAP) as a measure of Cognitive Fusion and Self-As Context in Students

1. Abstract.....	70
1.1 Background	70
1.2 Method	70
1.3 Results	70
1.4 Conclusion.....	70
2. Introduction	72
2.1 Overview of ACT and RFT.....	72
2.2 Cognitive Defusion and Self-As-Context Processes	73
2.3 Previous Research and Limitations	75
2.4 The IRAP	76
2.5 The Current Study	77
2.6 Aims and Hypothesis	77
3. Method.....	79
3.1 Power Calculation	79
3.2 Participants.....	79
3.2 Measures	80
3.2.1 The IRAP.....	80
3.2.2 Psychological Distress.....	83
3.2.3 Self-Esteem	83
3.2.4 Cognitive Fusion	83
3.2.5 Psychological Inflexibility	83
3.2.6 Social Desirability Bias	84
3.3 General Procedure.....	84
3.4 Data Preparation.....	85
3.5 D Scores	85
3.6 Data Analysis	86
4. Results	87
4.1 Attrition.....	87
4.2 IRAP Effects	88
4.3 Inverted IRAP Scores.....	90
4.4 Correlations between self-report measures (Hypothesis 1).....	91

4.5	Correlations among self-report <i>and D-IRAP</i> scores (Hypothesis 2)	91
4.6	Predictive validity of the IRAP (Hypothesis 3).....	92
4.6.1	Psychological Distress.....	93
4.6.2	Self-Esteem	93
5.	Discussion.....	95
5.1	Summary of Findings	95
5.2	Potential Confounding Variables of the IRAP Stimuli	96
5.3	Procedural Limitations	99
5.4	Nature of the Sample.....	99
5.5	Difficulties in Measuring ACT Processes	100
5.6	Conclusion.....	101
	References	102

Chapter III: Public Dissemination Report

1.	Introduction	111
2.	Literature Review:	112
2.1	Background	112
2.2	Method	112
2.3	Results	112
2.4	Conclusions	113
3.	Research Paper	114
3.1	Background	114
3.2	Method	114
3.3	Results	115
3.4	Conclusions	115
4.	Summary.....	116
	References	117

List of Appendices

Appendix 1. Ethical Approval Email	118
Appendix 2. Participant Information Sheet	119
Appendix 3. Informed Consent Form.....	123
Appendix 4. IRAP Instructions	124
Appendix 5. Depression Anxiety Stress Scale- 21	127
Appendix 6. Rosenberg Self-Esteem Scale	128
Appendix 7. Cognitive Fusion Questionnaire	129
Appendix 8. Acceptance and Action Questionnaire- Version II.....	130
Appendix 9. Marlowe Crowne Social Desirability Scale.....	131
Appendix 10. Screening Questions	132

List of Figures

CHAPTER I: The Effectiveness of Acceptance and Commitment Therapy with Children and Adolescents: A Meta-Analysis

Figure 1. Search Strategy Diagram.....	13
Figure 2. Forest Plot of Depression Post Intervention with and without Moghanloo et al., (2015)	33
Figure 3. Forest Plot of Depression Follow-up	35
Figure 4. Forest Plot of Anxiety Post Intervention.....	36
Figure 5. Forest Plot of Anxiety Follow-Up	38
Figure 6. Correct Funnel Plot for Anxiety Follow-Up	38
Figure 7. Forest Plot of Psychological Flexibility	39
Figure 8. Forest Plot of Functioning Self-Report Post Intervention.....	41
Figure 9. Forest Plot of Functioning Self-Report Follow-Up	41
Figure 10. Funnel Plot of Functioning Self-Report Follow-Up	42
Figure 11. Forest Plots with and without Martin et al (2015)	43

CHAPTER II: The Empirical research Report: Exploring the Implicit Relational Assessment Procedure (IRAP) as a measure of Cognitive Fusion and Self-As Context in Students

Figure 1. Defusion/Fusion and Self-As-Context Processes 74
Figure 2. Trial Types and Required Responses for Block 1 and 2..... 82
Figure 3. D-IRAP Scores..... 89
Figure 4. Inverted IRAP Trial-Types 90
Figure 5. Correlations between IRAP Trial-Types and Self-Report Measures..... 92

List of Tables

CHAPTER I: The Effectiveness of Acceptance and Commitment Therapy with Children and Adolescents: A Meta-Analysis

Table 1. Search Terms 10
Table 2. Meta-Analysis Inclusion/Exclusion Criteria 11

CHAPTER II: The Empirical research Report: Exploring the Implicit Relational Assessment Procedure (IRAP) as a measure of Cognitive Fusion and Self-As Context in Students

Table 1. Inclusion/Exclusion Criteria..... 80
Table 2. Target Statements and Target Words Used in the IRAP..... 81
Table 3. Descriptive Statistics of Completers and Non-completers of the IRAP. 88
Table 4. Hierarchical Regression for Psychological Distress..... 93
Table 5. Hierarchical Regression for Self-Esteem. 94

CHAPTER I: The Literature Review

**THE EFFECTIVENESS OF ACCEPTANCE AND COMMITMENT THERAPY WITH
CHILDREN AND ADOLESCENTS: A META-ANALYSIS**

1. Abstract

1.1 Background

There is growing interest in using Acceptance and Commitment Therapy (ACT) with children and adolescents. However, compared to the adult literature there is a lack of evidence demonstrating effectiveness with younger people. This is the first meta-analysis to investigate the effectiveness of ACT for children and adolescents.

1.2 Method

A systematic literature search yielded twenty studies (N=1,897) based on the inclusion/exclusion criterion. The current literature base is broad in focus and so the four most commonly reported outcomes (depression, anxiety, functioning and psychological flexibility) were extracted. A specifically designed quality criterion was developed to measure the risk of bias in each study.

1.3 Results

Small positive effects favouring ACT over control conditions were found for depression ($g=0.24$, 95% CI [0.01-0.42]), depression follow-up ($g=0.46$, 95% CI [0.06-0.86]), anxiety ($g=0.38$, 95% CI [0.05-0.72]), anxiety follow-up ($g=0.31$, 95% CI [0.04-0.57]) and psychological flexibility ($g=0.39$, 95% CI [0.08-0.70]). Whereas, moderate effects were produced for functioning ($g=0.71$, 95% CI [0.30-1.11]), which increased to large at follow-up ($g=0.87$, 95% CI [0.24-1.49]) and was consistent with parent reports of functioning ($g=0.85$, 95% CI [0.39-1.31]). A number of outcomes became non-significant when weighted for methodological quality.

1.4 Conclusion

Overall, the meta-analysis supports that ACT is effective with children and adolescents. However, this was based on a limited number of studies subject to various methodological limitations. Therefore, these results are tentative and further research is needed to draw firmer conclusions.

2. Introduction

An evidence-base of multiple meta-analysis demonstrates the efficacy of Acceptance and Commitment Therapy (ACT) within the adult population (A-Tjak et al., 2015; Hayes, Luoma, Bond, Masuda, & Lillis, 2006; Öst, 2014; Powers, Vörding, & Emmelkamp, 2009). These reviews report superior outcomes for ACT compared to control conditions including, waitlist controls, treatment as usual (TAU) and placebos across a range of disorders (e.g. anxiety disorder, weight problems and chronic pain). There is an expanding body of literature investigating the applicability and effectiveness of ACT with children and adolescents (Coyne, McHugh, & Martinez, 2011; Halliburton & Cooper, 2015; Swain, Hancock, Dixon, & Bowman, 2015). While this indicates that this approach shows promise, it is difficult to draw conclusions from the methodologically limited studies and diverse literature.

2.1 Overview of ACT

ACT is a third-wave mindfulness-based behavioural therapy that aims to increase psychological flexibility and enable individuals to behave according to their values (Levin, Hildebrandt, Lillis, & Hayes, 2012). Drawing upon Relational Frame Theory (RFT) (Hayes, Villatte, Levin, & Hildebrandt, 2011), ACT highlights linguistic and cognitive processes associated with psychological distress. It is thought that difficulties arise when a person's behaviour is guided by inflexible verbal representations of ongoing experience, rather than by direct experience of what is happening in the environment (Hayes et al., 2006).

ACT aims to help individuals respond more flexibly to their internal experiences by targeting six interrelated processes: Acceptance, Cognitive Defusion, Contact with Present Moment, Self-As-Context, Values and Committed Action (Luoma, Hayes, & Walser, 2007). While these processes are presented as distinct, it has been recognised that some of the processes may share some overlapping aspects such as cognitive defusion and self-as-context,

both involving distancing from thoughts. Importantly, while psychological flexibility is proposed to mediate increased wellbeing and decreased clinical symptoms (Ciarrochi, Bilich, & Godsell, 2010) symptom reduction is not the primary goal of ACT. Psychological flexibility is the primary goal and the Acceptance and Action: Second Version (AAQ-II) (Bond et al., 2011) has been developed as a unidimensional measure of this.

2.3 Rationale for Investigating the Effectiveness of ACT with Children and Adolescents

The prevalence rates of mental health disorders of children and adolescents have been reported as 10% for children and 20% for adolescents (Green, McGinnity, Meltzer, Ford, & Goodman, 2005). Early symptoms of psychological distress are predictive of mental illnesses in adulthood (Keenan, Feng, Hipwell, & Klostermann, 2009) and 75% of mental health problems are established by the age of 24 (Kessler et al., 2005).

The most researched intervention, Cognitive-Behavioural Therapy (CBT), has shown promising results however, a significant number of children have not responded to CBT. One in four children do not benefit from CBT (Ollendick, Öst, Reuterskiöld, & Costa, 2010) and 40% of children who receive CBT for an anxiety disorder still meet criteria for one post-intervention (Cartwright-Hatton, Roberts, Chitsabesan, Fothergill, & Harrington, 2004; James, Soler, & Weatherall, 2005). Therefore, there is a need to examine the effectiveness of a range of alternative treatments in addition to CBT.

Given that language is thought to be the core process underlying psychological distress in adults (Hayes, Pistorello & Levin, 2012), similar underlying processes of distress are expected in children and adolescents. Research demonstrates a strong relationship between high psychological inflexibility and symptoms of anxiety, depression, ADHD and

affective problems in adolescents (Greco & Hayes, 2008; Venta, Sharp, & Hart, 2012).

Therefore, ACT may demonstrate similar levels of effectiveness in children/adolescents as adults.

The ACT emphasis on teaching generic psychological skills (Lundgren, Dahl, & Hayes, 2008) suggests its potential use with non-clinical populations to reduce risk of developing mental health disorders. This may be particularly salient for younger individuals, who have had less time to develop more entrenched patterns of experiential avoidance (Greco, Blackledge, Coyne, & Ehrenreich, 2005). Prevention programmes are of increasing interest and have been outlined as a priority in the Five Year Forward Plan (NHS England, 2015). Thus, a statistical review of ACT with children/adolescents is timely.

2.4 Previous Reviews looking at ACT for Children and Adolescents

Four existing reviews of ACT outcome studies with children and/or adolescents across a range of conditions demonstrate early indications of effectiveness (Coyne et al., 2011; Halliburton & Cooper, 2015; Murrell & Scherbarth, 2006; Swain et al., 2015). One of these reviews was conducted systematically (Swain et al., 2015), which provided a synthesis of both published and unpublished studies. The other reviews examined the effectiveness of ACT for paediatric pain (Pielech, Vowles, & Wicksell, 2017), and chronic physical health conditions (Wicksell, Kanstrup, Kemani, Holmström, & Olsson, 2015). Finally, a meta-analysis examined the effectiveness of mindfulness-based interventions inclusive of ACT interventions (Kallapiran et al., 2015) however, this review was not specific to ACT interventions and therefore it is difficult to draw generalised conclusions regarding the efficacy of ACT as a specific intervention.

Generally, the previous reviews have drawn tentative but favourable conclusions regarding the efficacy of ACT across a number of presenting difficulties. Although, in the mindfulness meta-analysis (Kallapiran et al., 2015) ACT was only found to be comparable with active treatment, however this was only based on one or two ACT studies. The reviews agreed that the literature is still in its infancy and the quality of the evidence varies considerably across trials (Swain et al., 2015). Methodological limitations include small sample sizes, poorly defined control conditions and a lack of randomisation.

2.5 Moderators of therapeutic efficacy

It is unclear what factors may influence the effectiveness of ACT for children and adolescents. The reviews agree that age could be an important factor to consider in delivering ACT, particularly if a child needed to reach a certain developmental stage. However, it has been highlighted that there is a significant lack of evidence in children under 11-years-old (Swain et al., 2015) and so there is little guidance. It is also unclear how biological and emotional changes associated with puberty may effect engagement and how this impacts recommendations about session length and intensity (Halliburton & Cooper, 2015; Pielech et al., 2017).

Parental involvement within the intervention may also act as a factor in outcome (Murrell & Scherbarth, 2006; Pielech et al., 2017; Wicksell et al., 2015). However, it is not clear under what circumstances and in what form parental involvement might be helpful.

2.6 Limitations of Previous Reviews

The previous reviews presented with a number of limitations including a lack of systematic rigour. While Swain et al (2015) were systematic in their approach, the inclusion criterion was broad and so, studies with varied methodologies were included such as case

studies that have limited generalisability. This review also included unpublished papers that were not subject to peer review and as such, the risk of bias may be difficult to assess in these studies. All of the reviews (excluding Kallapiran et al., 2015) provided a synthesis of the literature based upon what the individual studies found. This does not allow the studies to be considered in context of all other studies and it is unclear what level of measurement error and natural variation there may be across the whole data set. There is also limited evidence regarding how methodological quality may be impacting these outcomes, which is important given the methodological issues within the literature.

2.7 Aims

The current meta-analysis will seek to provide a statistical summary of the data from outcome studies for ACT with children and/or adolescents across a range of presenting difficulties. While there are only a small number of studies for each presenting difficulty, it is hoped that the review can help to elicit whether it is helpful as a transdiagnostic approach. Such a review will offer an update on the literature, given that the previous systematic review (Swain et al., 2015) only included studies up until December 2014. This review also aims to investigate potential variables that may be moderating the effect of an intervention, in order to provide more information as to what aspects of ACT may be helpful, for whom and under what conditions.

3. Method

3.1 Search Strategy

A systematic search of the literature was conducted until March 2018 using PsychINFO, PubMed and Web of Science databases, building upon and adding a quantitative synthesis of an existing review of the effectiveness of ACT with children (Swain et al., 2015). The search terms that were used to identify these two areas (ACT and children and/or adolescent) were then combined (See Table 1). The references of the included studies were also examined in order to identify any further studies that had not been detected by the electronic search.

The Association for Contextual Behavioural Science (ACBS) webpage (<http://contextualscience.org/>) was also searched. The ACBS is an online research community holding a list of publications relevant to the field. A search was conducted on all their publications using all search terms related to the “Children and/or adolescents” construct outlined in Table 1.

Table 1. Search Terms

Construct	Free Text Search Terms	Method of Search	Limits
Acceptance and Commitment Therapy	<p>“Acceptance and commitment therapy”</p> <p>“ACT”</p>		
Children and/or adolescents	<p>“child*”</p> <p>“teen*”</p> <p>“adolescen*”</p> <p>“school*”</p> <p>“infan*”</p>	<p>Free search terms</p> <p>All search terms combined with <i>OR</i></p>	<p>Peer reviewed articles</p> <p>1967-March 2018</p>

3.2 Inclusion/exclusion criteria

Full inclusion/exclusion criteria are described in Table 2. Due to the heterogeneous nature of the literature base and with it still being in its infancy the criterion was kept broad. This included being a controlled intervention study, with ACT as the primary intervention with children/adolescents. The only restriction placed on study design was to ensure that there were sufficient data to calculate an effect size. No restrictions were placed on disorder or problem of interest, setting, or type of control or timeframe or follow-up.

Table 2. Meta-Analysis Inclusion/Exclusion Criteria

Inclusion criteria	Justification
<p>Nature of ACT intervention:</p> <p>Studies that report interventions that have employed a minimum of two of the ACT processes, as set-out by Swain et al (2015). These may include adaptations of ACT and may be called acceptance-based interventions.</p>	<p>This is to ensure that the intervention studies are not based on a single technique described within the ACT model. For example, mindfulness could be considered one of the ACT processes. However, there have already been reviews conducted on this (Kallapiran et al., 2015) and the aim of this is to test the effectiveness of the interlinking processes used in ACT.</p>
<p>In order to reflect the broad landscape of the literature, the intervention may also incorporate other therapeutic elements.</p>	<p>This is to reflect the heterogeneous nature of the presenting populations and differing elements of treatment that are included with different client groups. For instance, many studies include family involvement and/ or things such as pain education. Combination with other therapeutic elements will be accounted for in the quality criteria.</p>
<p>The intervention needs to be delivered face-to-face, rather than by telephone or online.</p>	<p>To reduce heterogeneity of the treatment given in each study.</p>
<p>Participant focus</p> <p>Studies that target interventions at child and/or adolescent populations. This is to include studies up to the age of 21. This does not include parenting-focused interventions.</p>	<p>This is to address a gap in the literature of assessing ACT's effectiveness with children and adolescents. Many of the studies reporting on outcomes for adolescents report up to the age of 21. The endpoint of adolescence is increasingly considered as now being well into the 20s (Sawyer, Azzopardi, Wickremarathne & Patton, 2018).</p>
<p>Outcome data</p> <p>The studies are required to report either Means and Standard Deviations, or F- Test statistics, Cohen's <i>d</i> effect size or an <i>r</i> effect size.</p>	<p>To ensure that outcomes can be calculated into an effect size.</p>
Exclusion criteria	Justification
<p>Type of article</p> <p>The following article types were excluded: meta-analysis/theoretical papers/ reviews/commentaries/ clinical guidance/non-outcome focused studies i.e. longitudinal/association studies/case studies/validation of psychometric scales/qualitative papers</p>	<p>These articles do not provide the outcome data needed for this meta-analysis.</p>
<p>When the study does not present group data and only provides individual scores.</p>	<p>Also, requires some methodological control and reduces increased bias of having less than 10 participants.</p>

The systematic search yielded 566 articles once duplicates (n= 448) were removed (see Figure 1). Articles were screened by title and abstract. The three most common reasons for exclusion were: not being an ACT intervention (n=196), not focused on a child/adolescent population (n=133) and not being a treatment outcome study (i.e. looking at the association between processes, a review article or assessing the validation of a psychometric tool; n=168). The remaining 56 articles were then reviewed in more detail against the exclusion criteria. 16 articles were included based on the inclusion/exclusion criteria. Three additional articles were identified from the ACBS website. Thus, 19 articles satisfied criteria for the meta-analysis. One article included two separate trials and so 20 trials are included overall.

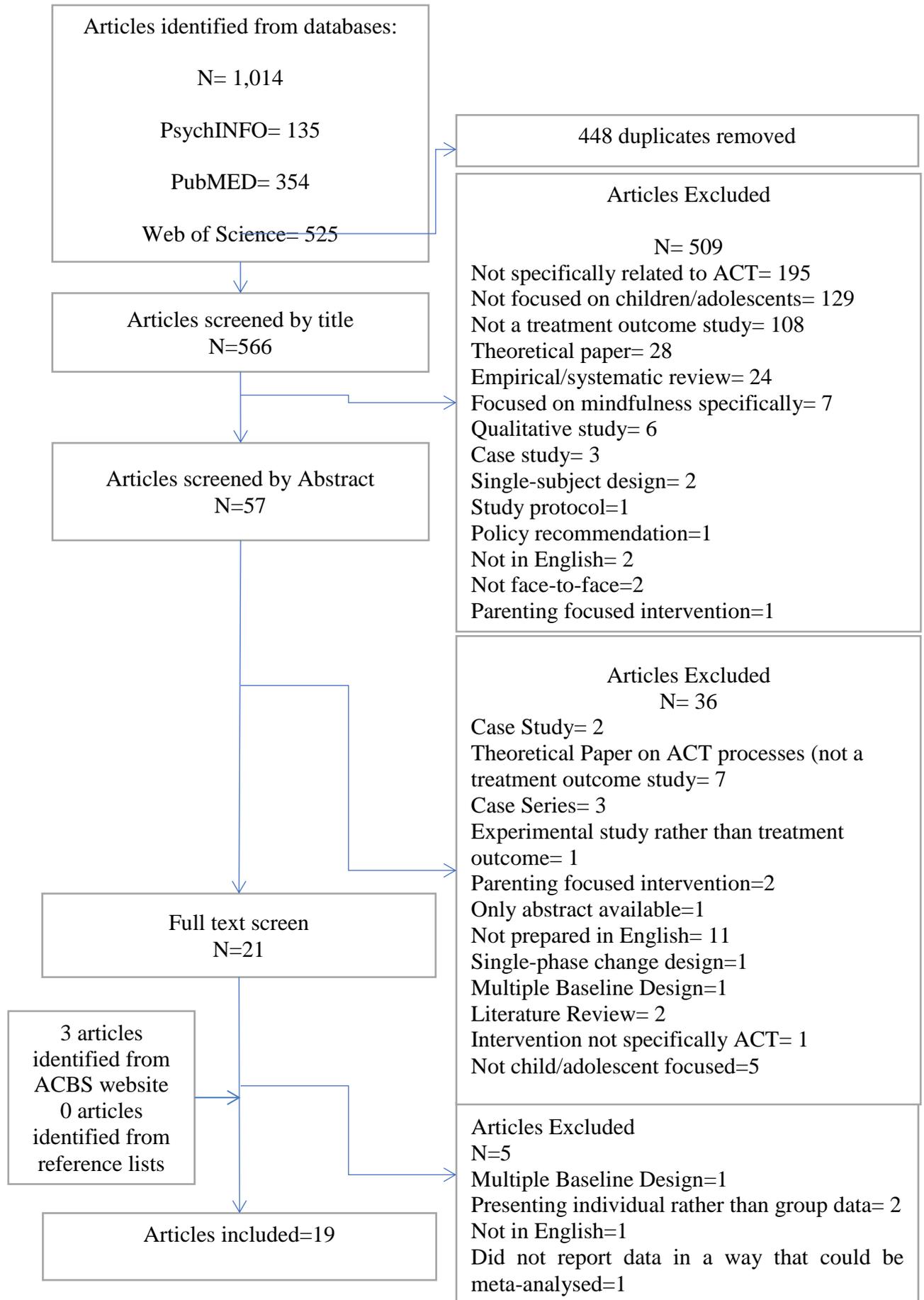


Figure 1. Search Strategy Diagram.

The key characteristics of the studies are presented in Table 3.

Table 3. Study Characteristics

Study	Sample Size	Age (range and mean)	% Female	Presenting problem	Country	Design	Problem Domain	Setting	Type of control	Intervention format	Parental involvement	Intervention length
<i>Azadeh, Kazemi-Zahrani, & Besharat (2015)</i>	30	15-16 15.43	100	Social anxiety	Iran	BG	Mental Health	School	No treatment	G	N	10 x 90 minutes Weekly
<i>Burckhardt, Manicavasagar, Batterham, & Hadzi-Pavlovic (2016)</i>	267	15-18 16.37	38	Prevention	Australia	RCT	Prevention	School	TAU ('pastoral classes')	G	N	16 x 30 minutes, twice weekly, over 3 months (9 sessions ACT)
<i>Burckhardt et al. (2017)</i>	48	14-16 15.64	42	Prevention	Australia	BG	Prevention	School	TAU ('pastoral classes')	G	N	4.6 hours of 25 minute workshops
<i>Gauntlett-Gilbert, Connell, Clinch, & McCracken (2013)</i>	98	10-19 15.6	75	Pain	UK	WG	Physical Health	Community	Pre/post	G	Y	90 hours over 15 days
<i>Ghomian & Shairi (2014)</i>	20	7-12 10.6	45	Pain	Iran	BG	Physical Health	Community	Unspec	Unspec	Unspec	Unspec
<i>Hancock et al. (2016)</i>	193	7-17 11.2	58	Anxiety	Australia	RCT	Mental Health	Community	WLC & CBT	G	Y	10 x 1.5 hours

Study	Sample Size	Age (range and mean)	% Female	Presenting problem	Country	Design	Problem Domain	Setting	Type of control	Intervention format	Parental involvement	Intervention length
<i>Hayes, Boyd, & Sewell (2011)</i>	38	12-18 14.9	71	Depression	Australia	RCT	Mental Health	Community	TAU (CBT)	I	N	Unspec
<i>Kanstrup et al. (2016)</i>	48	14-18 16	80	Pain	Sweden	BG (treated as WG)	Physical Health	Community	Individual/group	I & G	Y	18 sessions (2 hour –group 45 minutes individual)
<i>Livheim et al. (2015)-Australia</i>	66	12-18 14.6	87.88	Depression	Australia	BG	Mental Health	School	TAU (support from school nurse)	G	N	8 x 1.5 hours
<i>Livheim et al. (2015) Sweden</i>	32	12-18 Unspec	71.9	Stress	Sweden	RCT	Mental Health	School	TAU (support from school nurse)	G	N	6 x 15 hours
<i>Martin et al. (2016)</i>	14	12-20 16.9	70	Pain	USA	WG	Physical Health	Community	Pre/post	G	Y	3 x 2 hour sessions
<i>Moazzezi, Moghanloo, Moghanloo, & Pishvaei (2015)</i>	32	7-15 10.58	30.56	Diabetes	Iran	BG	Physical Health	Community	TAU	G	N	10 x 90 minutes
<i>Moghanloo, Moghanloo, & Moazezi (2015)</i>	34	7-15 10.47	50	Diabetes	Iran	BG	Physical Health	Community	TAU	G	N	10 x 90 minutes

Study	Sample Size	Age (range and mean)	% Female	Presenting problem	Country	Design	Problem Domain	Setting	Type of control	Intervention format	Parental involvement	Intervention length
<i>Pahnke, Lundgren, Hursti, & Hirvikoski (2014)</i>	28	13-21 16.5	25	High functioning Autism Spectrum Disorder	Sweden	BG	Prevention	School	WLC	G	N	2 x 40 minutes weekly & 6-12 minutes daily mindfulness
<i>Rostami, Veisi, Dehkordi, & Alkasir (2014)</i>	40	12-16 Unspec	0	Social anxiety in learning disability	Iran	BG	Prevention	School	TAU	G	N	10 x 1 hour
<i>Theodore-Oklota, Orsillo, Lee, & Vernig (2014)</i>	210	Unspec 12.45	44	Relational aggression	USA	BG	Prevention	School	TAU	G	N	3 x 48 minutes
<i>Timko, Zucker, Herbert, Rodriguez, & Merwin (2015)</i>	47	14-21 Unspec	89	Anorexia Nervosa	USA	WG	Mental Health	Community	Pre/post	G	Y	20 sessions over 24 weeks
<i>Van Der Gucht et al (2016)</i>	616	14-21	-	None	Belgium	RCT	Prevention	School	TAU	G	N	4 x 120 minutes
<i>Wicksell, Melin, & Olsson (2007)</i>	14	13-20 17	79	Pain	Sweden	WG	Physical Health	Community	Pre/post	I	Y	Av 14.4 hours child 2.4-hours parent
<i>Wicksell, Melin, Lekander, & Olsson (2009)</i>	32	10-18 14.8	78	Pain	Sweden	RCT	Physical Health	Community	TAU	I	Y	10 z 1 hour 1-2 hour parent group

Note: RCT= Randomised Control Trial; BG= Between-group; WG= Within-group; CO= Clinical Outpatient; CI= Clinical Inpatient; NC= Non-clinical; G= Group; I= Individual; TAU= Treatment as Usual; WLC= Waitlist Control; CBT= Cognitive-Behavioural Therapy; N= No; Y= Yes; Unspec= unspecified.

3.3 Quality Criteria

Quality criteria were developed to assess any risk of bias within the studies. The quality criteria were adapted from existing frameworks including: Downs & Black, (1998) The Cochrane Collaboration Risk of Bias Tool (Higgins et al., 2011) and the Risk of Bias Assessment Tool for Nonrandomised Studies (Kim et al., 2013). The framework assessed risk of bias in eight domains: Methodological Bias, Measurement of Outcomes, Blinding of Outcome Assessors, Attrition Bias, Selective Reporting, Treatment Fidelity, Treatment Validity and Power (see Table 4). Each domain was rated as either “Low”, “Unclear” or “High” risk of bias.

Table 4. Quality Criteria

Domain	Details	Risk of Bias
Methodological Bias	<p>What is the study design and the type of control used within the study?</p> <p>If using randomisation, have they described the method of allocation clearly? Has this allowed for the production of comparable groups?</p>	<p>High Risk-Within-group studies.</p> <p>Unclear Risk-Between group/quasi-randomised and randomisation studies where methods of allocation have not been clearly described.</p> <p>Low Risk-Randomised Control Study, where procedures of randomisation have clearly been described and allocation has been concealed.</p>
Measurement of Outcomes	<p>Is outcome measure valid and reliable for this population?</p> <p>Have they used an ACT specific outcome?</p> <p>Is the outcome measure using full scale or subscales?</p>	<p>High Risk- Only reported un-validated outcome measure/ used a subscale which does not report reliability/validity.</p> <p>Unclear Risk- No report of an ACT specific outcome measure.</p> <p>Low Risk-Reliable and valid outcome measures and an ACT specific measure is used.</p>
Blind Outcome Assessment	<p>Are the outcome assessors blind to participant allocation?</p>	<p>High Risk- No blinding of outcome assessment has taken place.</p> <p>Unclear Risk- Not reported on.</p> <p>Low Risk- Outcome assessment is blind.</p>
Selective Reporting	<p>Is there evidence of selective outcome reporting? i.e. only significant results reported.</p> <p>Are there measures that have not been reported in the results that have been mentioned in the method section?</p>	<p>High Risk- Not reported full outcome measures that are stated in the method section/ reported only a subsample of results/only significant results.</p> <p>Unclear Risk- Not all descriptive statistics are presented.</p>

Domain	Details	Risk of Bias
		Low Risk- Reported all results of measures as outlined in the method.
Attrition Bias	<p>Is there incomplete data due to attrition?</p> <p>Has this been handled appropriately?</p>	<p>High Risk- Completer only analysis (greater than 30% attrition).</p> <p>Unclear Risk – No report of attrition/or between 20-30% attrition.</p> <p>Low Risk – low risk of bias (intention to treat analysis or completer analysis with >80% of sample)/20-30% attrition with adequate control.</p>
Treatment Fidelity	<p>Has treatment fidelity been assessed by recording of sessions or by supervision or screened for protocol adherence?</p> <p>What did the results of this show?</p>	<p>High Risk – No mention of treatment fidelity tests or processes used to ensure fidelity.</p> <p>Unclear Risk – Treatment fidelity undertook but not described/evaluated.</p> <p>Low Risk - Treatment fidelity described and adequate adherence to the model demonstrated.</p>
Treatment Validity	<p>Does the intervention follow a standardised/structured protocol manual?</p> <p>Is it combined with other treatments?</p> <p>Is it conducted by someone who is adequately trained to deliver the intervention?</p>	<p>High Risk- Combined with another treatment, no protocol.</p> <p>Unclear Risk- unclear if following protocol, training of those delivering the intervention not reported.</p> <p>Low Risk- Standalone ACT, based upon protocol and conducted by someone with suitable experience.</p>
Power	<p>Is there sufficient power to calculate an intervention effect?</p> <p>Small = <20 per arm</p> <p>Adequate > 20 per arm</p>	<p>High risk – Small sample with or without idiosyncratic feature (<20 per group).</p> <p>Unclear risk - Sufficient sample for generalisation but with some idiosyncratic feature (> 20 per group).</p> <p>Low risk- Sufficient sample for generalisation and representative of target population (>20 per group)</p>

A summary of the risk of bias for each study is presented in Figure 2.

Study	Methodological bias	Measurement of Outcomes	Blinding of Outcome Assessors	Selective Reporting	Treatment Fidelity	Treatment Validity	Attrition Bias	Power
Azadeh et al (2015)	Amber	Amber	Amber	Green	Red	Amber	Amber	Red
Burckhardt et al (2017)	Amber	Amber	Amber	Green	Amber	Green	Green	Red
Burckhardt et al (2016)	Amber	Amber	Amber	Red	Green	Red	Green	Amber
Gauntlett-Gilbert et al (2013)	Red	Green	Amber	Red	Green	Amber	Green	Amber
Ghomian & Shairi (2014)	Amber	Amber	Amber	Green	Red	Amber	Amber	Red
Hancock et al (2016)	Green	Green	Amber	Amber	Green	Green	Green	Green
Hayes et al (2011)	Green	Amber	Amber	Green	Green	Green	Green	Red
Kanstrup et al (2016)	Red	Amber	Amber	Green	Green	Amber	Amber	Red
Livheim et al (2015) Australia	Amber	Green	Amber	Green	Amber	Green	Green	Red
Livheim et al (2015) Sweden	Green	Green	Amber	Green	Red	Amber	Green	Red
Martin et al (2016)	Red	Amber	Amber	Green	Red	Amber	Green	Red
Moazzezi et al (2015)	Amber	Amber	Amber	Green	Red	Amber	Amber	Red
Moghanloo et al (2015)	Amber	Amber	Green	Green	Red	Green	Green	Red
Pahnke et al (2014)	Amber	Amber	Amber	Amber	Amber	Green	Green	Red
Rostami et al (2014)	Amber	Red	Amber	Red	Red	Amber	Amber	Red
Theodore-Oklota et al (2014)	Amber	Red	Amber	Green	Red	Green	Green	Green
Timko et al (2015)	Red	Green	Amber	Green	Green	Red	Red	Amber
Van Der Gucht et al (2016)	Green	Green	Amber	Green	Amber	Red	Green	Green
Wicksell et al. (2007)	Red	Green	Amber	Green	Red	Green	Green	Red
Wicksell et al (2009)	Green	Red	Green	Green	Green	Red	Green	Red

Figure 2. Summary of Applied Quality Criteria.

Red indicates high risk of bias, amber marks an unclear risk of bias and green is a low risk of bias.

3.3.1 Methodological Bias

Methodological bias was rated as low risk in five studies. Ten studies were rated as unclear risk, which included those regarded as between-group designs or ‘quasi’-randomised designs (i.e., Burckhardt et al., 2017; Pahnke et al., 2014) or using ‘cluster randomisation’ (Burckhardt et al., 2016; Theodore-Oklotka et al., 2014). The other studies rated as unclear risk provided limited information on how the randomisation process was conducted (Azadeh et al., 2015; Ghomian & Shairi, 2014; Moazzezi et al., 2015; Moghanloo et al., 2015; Rostami et al., 2014). All within-subject design studies were rated as a high risk of bias.

3.3.2 Measurement of Outcomes

The majority of the studies used outcome measures that were valid and reliable. However, numerous studies were rated as unclear risk for not including an ACT specific outcome measure such as, psychological flexibility (Burckhardt et al., 2016, 2017; Ghomian & Shairi, 2014; Hayes et al., 2011; Moazzezi et al., 2015; Moghanloo et al., 2015; Pahnke et al., 2014). This was considered important given the primary goal of ACT is to improve psychological flexibility. Studies that included adapted versions of scales/subscales without reporting of reliability/validity were considered high risk (Azadeh et al., 2015; Martin et al., 2016; Rostami et al., 2014; Theodore-Oklotka et al., 2014; Wicksell et al., 2009)

3.3.3 Blinding of Outcome Assessors

The majority of studies did not report if those administering the outcome assessments were blinded. Only two studies reported using blind outcome assessors (Moghanloo et al., 2015; Wicksell et al., 2009).

3.3.4 Selective Reporting

Full reporting was generally considered good across studies. However, there were some studies that did not report some of the subscales that were mentioned in the method (Gauntlett-Gilbert et al., 2013; Rostami et al., 2014). One study reported results from a subset of the sample (Burckhardt et al., 2016) and two studies did not report means and standard deviations (Kanstrup et al., 2016; Pahnke et al., 2014).

3.3.5 Treatment Fidelity

Only two of the studies reported measurement and results of treatment fidelity (Burckhardt et al., 2016; Hancock et al., 2016). However, the rest of the studies were considered low risk due to having adequate procedures in place to maintain treatment fidelity such as having an external supervisor (Hayes et al., 2011) and having audiotapes independently reviewed (Timko et al., 2015). If studies did not report any procedures to ensure treatment fidelity, they were regarded as being high risk.

3.3.6 Treatment Validity

The majority of the studies included clear protocols for the intervention that appeared to be grounded in at least four of the six ACT processes and suitably referenced the ACT literature. Threats to treatment validity arose from studies combining ACT with other approaches such as, Positive Psychology (Burckhardt et al., 2016), Family Therapy (Timko et al., 2015) and exposure (Wicksell et al., 2009). In some studies, it was unclear who delivered the intervention and whether they were suitably trained in ACT (Azadeh et al., 2015; Gauntlett-Gilbert et al., 2013; Ghomian & Shairi, 2014; Moazzezi et al., 2015). In one study, teachers with only two days of ACT training delivered the intervention and due to their relative inexperience were considered as high risk of bias to treatment validity (Van Der Gucht et al., 2016).

3.3.7 Attrition Bias

Attrition rates were variable however the studies were consistent at providing reasons for drop-outs and providing adequate controls for this, including intention-to-treat analyses and carrying out statistical analyses between completers and non-completers. A number of studies did not report attrition data and while this may have been because they did not have any drop-outs this could not be assumed (Azadeh et al., 2015; Ghomian & Shairi, 2014; Moazzezi et al., 2015; Rostami et al., 2014). One study reported attrition above 30% and while they controlled for this in their analysis, the author was using end of intervention data rather than change data and so this was regarded as a high risk of bias for the current review (Timko et al., 2015).

3.3.8 Power

The small sample sizes were the largest risk of bias amongst studies, with 14 of the studies having sample sizes of 20 or less per group. Only one study reported on power, stating that in order to obtain a power of 0.85, they would have needed to allocate 75 to each condition, whereby only 32 (ACT) and 26 (TAU) were allocated (Livheim et al., 2015 - Sweden). Also, when some studies did have more than 20 participants per group they were sometimes subject to bias due to the sample having idiosyncratic features. This included participants all being from a private independent high school who were in the top quartile of socio-economic advantage (Burckhardt et al., 2016) or samples being predominantly female (Gauntlett-Gilbert et al., 2013; Timko et al., 2015).

3.3.9 Summary

Overall, the levels of bias varied across studies. There was only one reported study that did not report any high risk of bias in any domain (Hancock et al., 2016). Notably, there was particularly high risk of bias across studies in the domain of Power. Treatment Fidelity

was deemed as an area of high risk whereby it was unclear what measures were in place to address this. Studies with medium-to-high risk of bias were included due to the low number of studies in this field. Therefore, results of this meta-analysis should be interpreted with caution until findings are replicated by future trials with improved methodology and larger sample sizes.

3.4 Data Extraction

The four most commonly reported outcomes were used in the analysis to gain an overview as to whether ACT is effective for children/adolescents. The outcome domains included: Depression, Anxiety, Psychological Flexibility and Functioning. Functioning related to any activity limitation based upon physical or mental health difficulties. The extracted outcome measures are presented in Table 5.

As a result of the outcomes being measured by different scales, the analysis could not be carried out on the raw difference in means. In order to create a common metric, the mean difference between the ACT and control group and between pre-and post-measures in within-group studies was divided by the pooled standard deviation in a study to produce a standardised mean difference (SMD). The SMD has been found to slightly overestimate the true effect size, especially in smaller samples (Borenstein et al., 2009). Accordingly, a correction for this bias (Hedges, 1981) was used for all calculations to remove this bias. For ease of interpretation, the transformed effects size (Hedge's g) was back-transformed to the more familiar Cohen's d (SMD) in the summary tables and forest plots. The magnitude of summary effects will be interpreted by Cohen's categorical levels of small (0.2), medium (0.5) and large (0.8; Cohen, 1988).

Three of the studies did not provide their data in an appropriate format. Hayes et al (2011) reported standard errors, which had to be converted into standard deviations prior to analysis. Pahnke et al (2014) presented some of their outcomes in graphical format. Therefore, the means and standard deviations were calculated from the confidence intervals provided on the graph and the sample size, and are therefore only approximations. Finally, Kanstrup et al (2016) provided an effect size in terms of a correlation (r), which was then converted to a standardised mean difference (d) using the calculation provided by Borenstein et al (2009).

If a study reported more than one measure for each domain, a decision was made about which measure to use. This was to ensure that only one effect size from each study was reported for each outcome category, to avoid artificially inflating the sample-size of the meta-analysis. Validated self-report measures were preferred over self-reported frequency data and self-report ratings were favoured over parent/teacher ratings. Also, measures intended for use as an outcome measure were favoured over diagnostic measures that were included for screening purposes. If two measures were reported for the same domain, then the most consistently used measure across studies would be used and if more than one measure still remained, then the more reliable measure would be chosen. If multiple measures were included for one outcome but measured different aspects of an outcome for example, social functioning versus physical functioning then the aspect that was more explicitly linked to the intervention aims was chosen.

Table 5. Extracted Outcome Data

Study	Depression	Anxiety	Functioning	Psychological Flexibility	Follow-up
Azadeh et al (2015)				AAQ-II ¹	None
Burckhardt et al (2016)	DASS ² -depression	DASS-anxiety			
Burckhardt et al (2017)	DASS-depression	DASS-anxiety			5 months
Gauntlett-Gilbert et al (2013)	BAPQ ³ -depression	BAPQ-pain specific anxiety	BAPQ- physical disability		3 months
Ghomian and Shairi (2014)			FDI ⁴ (child and parent)		1.5 months
Hancock et al (2016)		MASC ⁵	Psychosocial health related QOL ⁶	AFQ-Y ⁷	None (only vs CBT)
Hayes et al (2011)	RADS-2 ⁸				3 months
Kanstrup et al (2016)	CES-D ⁹		FDI (parent)	PIPS ¹⁰	None
Livheim et al (2015)-Australia	RADS-2			AFQ-Y8 ¹¹	None
Livheim et al (2015)-Sweden	DASS depression	DASS Anxiety		AFQ-Y17 ¹²	None
Martin et al (2015)	CES-D	PASS-20 ¹³	FDI (child and parent)		None
Moazezzi et al (2015)					None
Moghanloo et al (2015)	RCDS ¹⁴				None
Pahnke et al (2014)	BYI ¹⁵ depression	BYI anxiety			2 months
Rostami et al (2014)		Social anxiety scale			None
Theodore-Oklota et al (2014)				AFQ-Y	None
Timko et al (2015)				AFQ	3 months
Van der Gucht et al (2016)		YSR ¹⁶ Anxiety		AFQ-Y	12 months
Wicksell et al (2007)			FDI (child)		3 & 6 months
Wicksell et al (2009)	CES-D		FDI (child and parent)		3.5 & 6.5 months

¹ Acceptance and Action Questionnaire Version 2 ²Depression Anxiety Stress Scale ³Bath Adolescent Pain Questionnaire ⁴Functional Disability Inventory ⁵Multi-Dimensional Anxiety Scale for Children ⁶Quality of Life ⁷Avoidance & Fusion Questionnaire for Youth ⁸Reynolds Adolescent Depression Scale Version 2 ⁹Center for Epidemiological Studies Depression Scale ¹⁰Psychological Inflexibility in Pain Scale ¹¹Avoidance & Fusion Questionnaire for Youth 8 items scale ¹²Avoidance & Fusion Questionnaire for Youth 17 items scale ¹³Pain Anxiety Symptoms Scale ¹⁴Reynolds Child Depression Scale ¹⁵Beck Youth Inventories ¹⁶Youth Self Report

3.5 Data Analysis

The “Meta” package in R (R Core Team, 2015; Schwarzer, 2007; Schwarzer, Carpenter, & Rücker, 2015) was used to carry out a separate meta-analysis for each outcome category. As some of the studies carried out follow-up analyses, separate analyses were carried out for each time category to identify if there were any significant differences in effect over time. This was split into end-of-intervention data (at the end of the intervention or as close to the end point as possible) and short-term follow-up (2 months to 6 months following the intervention). Some of the studies also included the same functioning outcome measure but in different formats. They included both self-report and parent measures, which were split into separate analyses.

3.6 Meta-analytic model

A random effects model was used to calculate the meta-analytic effect for each outcome, in order to provide a mean of a distribution of effects, based on the assumption that true effect size will vary across studies. This model was selected over the fixed effects model, which assumes that the true effect size is identical across studies and that differences across studies is only due to sampling error (Borenstein et al., 2009). The assumptions underlying this model are rarely met within psychological research, where there are likely to be a number of factors which influence treatment efficacy. This is certainly true of the current review where it is likely that factors such as participant idiosyncrasies, methodological variations and discrepancies in how the intervention was delivered and measured will influence effect size. As a result of the random effects model accounting for various sources of bias, this will be reflected in larger Confidence Intervals (CI).

3.7 Quality Effects

In what could be considered an extension of the random effects model, the quality effects model was conducted. Where the random effects model views precision of effect size as a function of sample size, the quality effects model assumes that methodological quality will also influence effect size. Thus, this model gives more weight to those studies that demonstrate lower risk of bias and can be used to investigate whether the outcomes were sensitive to risk of bias in the primary studies (Doi & Thalib, 2008).

3.8 Quantifying Heterogeneity

The random effects model recognises that effect sizes may vary across studies as a result of uncontrolled factors. Thus, it is important to identify and quantify this heterogeneity. Heterogeneity was calculated for all outcomes using Cochrane's Q , Higgins I^2 and Tau. High levels of heterogeneity would suggest that factors other than the response to the ACT intervention may be influencing outcomes.

Borenstein et al., (2009) recommends the use of Higgins I^2 to decide on whether to investigate this further. Higgins et al., (2011) have put forward tentative benchmarks of 25%, 50% and 75%, as low, moderate and high respectively. Given the variability in study design and differences in presenting difficulties it was decided that a Higgins I^2 benchmark of 75% would be considered as 'problematic heterogeneity'.

3.9 Identifying Influential Studies

3.9.1 One Left Out Procedure

A "one left out" procedure was conducted to investigate whether any particular study was having a disproportionate influence on the overall meta-analytic effect. This involved systematically removing each study and recalculating the overall meta-analytic effect. If

removing a study resulted in a meta-analytic effect lying outside the original 95% CI this was considered disproportionately influential and the study will be omitted.

3.9.2 Influential Case Diagnostics

A series of influential case diagnostics¹ were also conducted to test whether a primary study was reporting effects consistent with other studies for an outcome. The following ‘leave-one-out diagnostics’ were obtained; externally standardised residual, DFFITS value, Cook’s distances, covariance ratio, the leave-one-out amount of residual heterogeneity, the leave-one-out test statistic for heterogeneity, externally standardised residual and DFBETAS values. A more detailed review of these diagnostics is provided by Viechtbauer & Cheung, (2010).

Accordingly, a primary study would be considered ‘influential’ if omission of that study would change the outcome of the meta-analysis and ‘inconsistent’ if that study does not report effects of same direction and/or magnitude of effect as the consensus within the literature. Influential studies will be omitted from the meta-analysis and influential studies will be highlighted in the reporting of the meta-analytic summary².

¹A study will be considered “inconsistent” if any one of the following is true: the absolute DFFITS value is larger than $3 \cdot \sqrt{p/(k-p)}$, where p is the number of model coefficients and K is the number of studies; The lower tail area of the chi-square distribution with p degrees of freedom is cut off by the Cook’s distance, which is larger than 50%; the hat value is larger than $3(p/k)$; Any DFBETAS value is larger than 1

² The procedure for identifying primary studies that were inconsistent with meta-analytic effect and deciding whether to omit them from analysis will be described. Firstly, if the meta-analysis was viewed as having problematic heterogeneity (>75%), the “one left out” procedure will be conducted. If the study is found to be inconsistent, this was removed from the analysis. If there was inadequate heterogeneity and no one study was found to be influential from this procedure, the “leave one out diagnostics” was reviewed. If any significant findings emerged, the study was not omitted, but the differential effect sizes were reported so that the reader was aware of potential biases.

3.10 Publication Bias

Publication bias was identified using visual examination of funnel plots. However, given the subjective nature of method (Terrin, Schmid, & Lau, 2005) a statistical test of funnel plot asymmetry (Egger, Davey Smith, Schneider, & Minder, 1997) was also conducted when there were ten or more studies. If publication bias was identified, a trim and fill procedure (Duval & Tweedie, 2000) and fail-safe N was calculated (Rosenthal, 1979). The trim and fill procedure firstly removes the most extreme small studies and then uses funnel asymmetry to estimate missing small null studies. The meta-analytic effect is then recalculated with these estimated studies included. In contrast, the fail-safe N is an estimation of the number of missing null studies that are required for the effect to be no longer significant.

3.11 Moderator Analyses

As previously stated, there was variation in the participant groups being investigated, as well as the way in which the intervention was delivered. Subsequently, moderator analyses were conducted on each outcome to investigate the impact on meta-analytic effect. These variables included: study design (RCT/between-groups/within-groups), intervention format (group/individual), problem domain (physical health/mental health/prevention), study setting (school/community), country of study (divided into Iran vs the rest of the world following pattern of results), parental involvement (yes/no), age, year of study, percentage of females and length of intervention. A subgroup analyses was conducted for categorical variables and a meta-regression was conducted on continuous variables.

4. Results

4.1 Participant Characteristics

Within the primary 20 studies (N=1,897) sample sizes ranged between 14 (Martin et al., 2016; Wicksell et al., 2007) to 616 (Van Der Gucht et al., 2016). Ages ranged from 7-21 years old, with the lowest mean age of 11 (Moghanloo et al., 2015) and highest of 17 (Wicksell et al., 2007). The proportion of males to females also varied greatly, with one study having an entirely female sample (Azadeh et al., 2015) and another having an all-male sample (Rostami et al., 2014). Participants were recruited from six different countries including Iran (n=5), Sweden (n=5), USA (n= 3), Australia (n= 5) and one from the UK and Belgium. Study designs included RCTs (n=6), between-group (n=9) and within-group (n=4) designs. The studies presented a range of presenting difficulties and were separated into three main areas: mental health (n=6), physical health (n=8) and prevention (n=6).

4.2 Intervention Characteristics

The ACT intervention mainly delivered in a group format (n=15) compared to individual sessions (n=3). One study compared group and individual formats (Kanstrup et al., 2016) and one that did not specify (Ghomian & Shairi, 2014). The length of the intervention varied from 20 sessions (Timko et al., 2015) to as little as three 48-minute sessions (Theodore-Oklota et al., 2014). One study combined ACT with Positive Psychology (Burckhardt et al., 2016) and a number of the studies included parental involvement (n=6). The most common control was TAU (n=13), with one study comparing ACT to a wait-list control group and Cognitive-Behavioural Therapy (Hancock et al., 2016).

4.3 Summary

Overall, significant positive associations were observed favouring ACT over control conditions, with the size of effects variable across outcome and time point (See Table 6).

Table 6. Summary of Meta-Analytic Effects.

	Number of Studies	Number of Participants	Random Effects model (* represents significance)		Heterogeneity			
			Summary Effect	95% CI	Higgins I ²	Tau ²	Cochrane's Q	Sig (p)
Depression Post Intervention	11	409	0.48*	0.03-0.93	83%	0.45	57.92	<.0001
Removal of extreme outlier studies	10	375	0.24*	0.01-0.42	32%	0.04	13.15	.016
Depression Follow-up	5	191	0.46*	0.06-0.86	46%	0.09	7.51	0.11
Anxiety Post Intervention	9	929	0.38*	0.05-0.72	80%	0.18	39.89	<.0001
Anxiety Follow-up	4	498	0.31*	0.04-0.57	41%	0.03	5.08	0.166
Psychological Flexibility Post Intervention	8	580	0.39*	0.08-0.70	79%	0.14	33.44	<.0001
Functioning (self-report) Post Intervention	6	307	0.71*	0.30-1.11	67%	0.15	7.93	0.009
Functioning (self-report) Follow-up	4	136	0.87*	0.24-1.49	71%	0.27	10.32	0.02
Functioning (parent report) Post Intervention	4	78	0.61	-0.02-1.25	62%	0.26	7.93	0.04
Removal of extreme outlier studies	3	58	0.85*	0.39-1.31	12%	0.02	2.72	0.32

NOTE: The magnitude of summary effects can be interpreted as small = 0.2, medium = 0.5 and large = 0.8

Grey shading denotes follow up.

4.4 Depression

Eleven of the studies included an outcome measure of depression (n=409). Of these eleven, five of them measured this outcome during a follow-up period (n=191), an average of 3.3 months following the end of the intervention.

4.4.1 Post Intervention

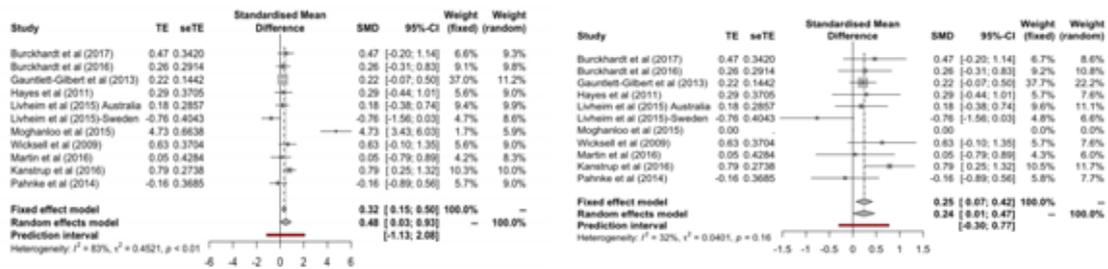


Figure 2. Forest Plot of Depression Post Intervention with and without Moghanloo et al., (2015)

The random-effects model produced a small, significant effect (SMD= 0.48, CI; [0.03-0.93]) in favour of ACT reducing depressive symptomology. However, problematic heterogeneity was observed ($I^2=82$), suggesting that there may be uncontrolled sources of variance influencing meta-analytic effect that require further investigation. As seen in Figure 2, Moghanloo et al (2015) is demonstrating a disproportionately larger effect size compared to the rest of the literature. This was confirmed using the “one left out” procedure, whereby omitting Moghanloo et al (2015) was found to produce an influential effect size, with its 95% CI falling outside the meta-analytic summary. Removal of this study also reduced

heterogeneity to an acceptable level ($I^2 = 32\%$). Consequently, this study was omitted (see Figure 2).

Removal of this study resulted in a small significant meta-analytic effect (SMD=0.24, CI [0.01-0.47]), favouring ACT reducing depression compared to control groups. However, this result lay just within the boundary of significance and may easily become non-significant as a result of minor changes within this literature. When weighted for methodological quality, the estimated effect reduced by 0.03 and became non-significant (SMD= 0.2, CI [-0.43-0.45]). Thus, methodological variation accounted for approximately 12% of the random effects estimate. Visual inspection of the funnel plot did not identify the presence of publication bias and Egger's test of asymmetry was non-significant ($P=0.66$).

Overall, the literature suggests a small significant effect of ACT reducing depression. However, the estimation of this effect is tentative given that the effect is on the boundaries of significance.

4.4.3 Follow-up

The random effects model indicated that the small significant effect observed for depression was maintained and consolidated at follow-up, see Figure 3 (SMD=0.46, CI [0.06-0.86]). Heterogeneity was deemed acceptable ($I^2=46\%$) and no further tests to identify sources of variance were required. Using the "one left out" procedure, no study was observed to exert disproportionately influence.

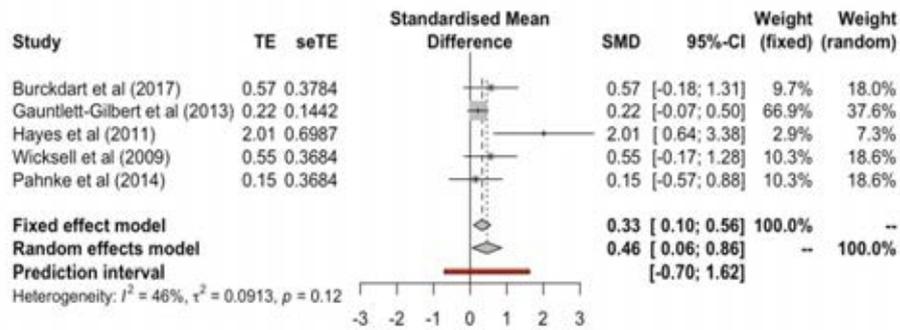


Figure 3. Forest Plot of Depression Follow-up

When weighted for methodological quality, this estimated effect increased substantially by 0.33 (SMD= 0.79, CI [0.28-1.29]). This suggests that the effect size would increase if the methodological quality across all studies matched the highest reported study quality in the analysis (Hayes et al., 2011). Investigations into publication bias were limited because of the small number of studies. However, visual inspection of the funnel plot did not indicate presence of publication bias and no studies were deemed to be missing using the trim and fill procedure.

In sum, ACT produced a small and significant reduction in self-reported depressive symptomology in children/adolescents at follow-up. The true effect size may even be larger, if all studies included were of higher methodological quality.

4.5 Anxiety

4.5.1 Post Intervention

Nine of the studies included an outcome measure of anxiety (n=929). Four of the nine studies included follow-up periods (n=498) an average of 5.5 months after the intervention ended.

As seen in Figure 4, the random effects model yielded a small significant effect (SMD= 0.38, CI [0.05-0.72]) in favour of ACT reducing self-reported anxiety. Heterogeneity was deemed unacceptably high ($I^2= 80\%$) and further testing is required to examine potential sources of variance.

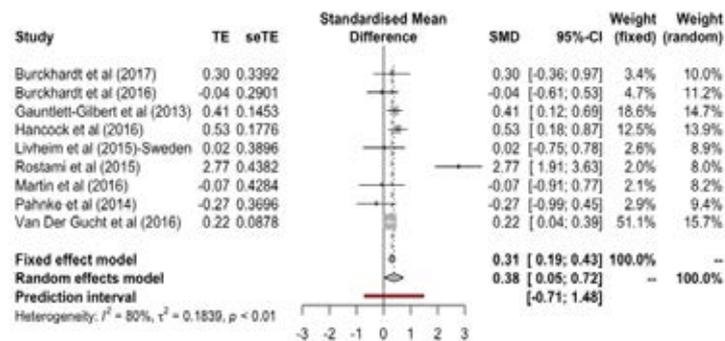


Figure 4. Forest Plot of Anxiety Post Intervention

Using the “one left out” procedure, no study was observed to be disproportionately influencing the meta-analytic summary. Rostami et al (2014) appeared to be exerting the biggest influence on overall effect, however its omission produced an effect that fell within the 95% CI. Nonetheless, this study was identified as ‘inconsistent’ by the “influential case diagnostics” and its omission significantly reduced heterogeneity ($I^2= 8$) and reduced the estimated summary effect to SMD= 0.26 (CI [0.12-0.40]). When studies were weighted for

methodological quality, there was an 18% decrease in overall effect size (SMD= 0.32, CI [-0.03-0.66]) and this effect was no longer significant. Thus, studies of lower methodological quality such as Rostami et al (2014), may be producing an overestimation of meta-analytic effect. Visual inspection of the funnel plot did not indicate the presence of publication bias, whereby there was a cluster of small studies in the bottom left, as supported by the trim and fill procedure.

In sum, a small positive effect favouring ACT reducing self-reported anxiety was reported. However, this is only a tentative conclusion given the sensitive nature of the meta-analytic effect and problematic heterogeneity observed. It was noted that one study may be exerting undue influence on the meta-analytic effect and thus, the estimated meta-analytic effect may fall somewhere in-between 0.26 to 0.38.

4.5.2 Follow-up

As seen in Figure 5, the random effects model demonstrated that the small significant effect for anxiety was maintained at follow-up (SMD=0.31, CI [0.04-0.57]). Heterogeneity was considered acceptable ($I^2= 41\%$). Using the “one left out” procedure, no study was identified as exerting disproportionate influence on meta-analytic effect. The quality effects model estimated a slight, but non-significant, increase in the effect of approximately 9% (SMD= 0.34, CI [0.01-0.66]).

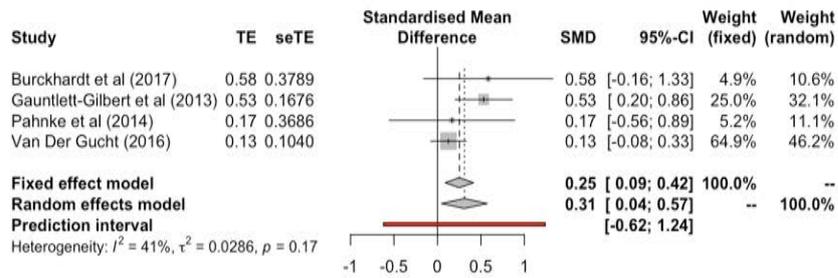


Figure 5. Forest Plot of Anxiety Follow-Up

Visual inspection of the funnel plot highlighted the possible presence of publication bias. Using the trim and fill procedure, one study was estimated to be added to the bottom left. The corrected funnel plot is presented in Figure 6. The corrected estimated effect size remained small and significant (SMD= 0.27, CI [0.04-0.50]). The fail-safe N was calculated to be 12, three times as many studies included suggesting that the result is robust to the effects of publication bias.

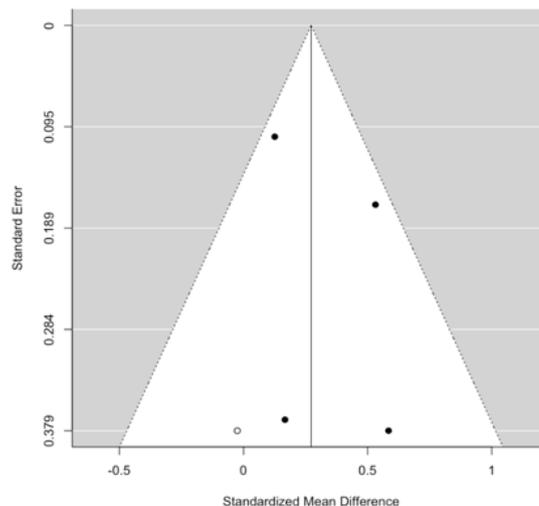


Figure 6. Correct Funnel Plot for Anxiety Follow-Up

Overall, a small significant effect was estimated in favour of ACT reducing self-reported anxiety at follow-up. However, the meta-analytic effect was bordering the significance boundary, with only minor variations in the literature required to lose significance. Thus, this result should be considered indicative.

4.6 Psychological Flexibility

Eight studies included a measure of psychological flexibility (n=580), with the majority of studies using an adaptation of the Action and Acceptance Questionnaire. As seen in Figure 7, the random effects model yielded a small, significant effect (SMD= 0.39, 95% CI [0.08-0.70]) of ACT improving psychological flexibility. Heterogeneity was found to be problematic ($I^2= 79%$) and further investigation into the sources of variation was required.

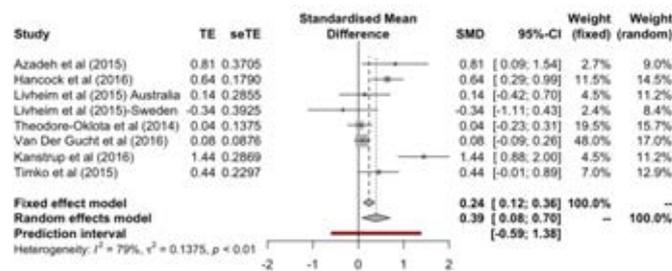


Figure 7. Forest Plot of Psychological Flexibility

The “one left out” analysis did not demonstrate that any study was disproportionately influencing effect size. The study exerting the largest influence (Kanstrup et al., 2016) was found to be inconsistent with the rest of the literature using the “influential case diagnostics”. Omitting this study would result in a smaller, but significant effect size in favour of ACT (SMD=0.25, CI [0.02-0.48]) and heterogeneity would reduce to an acceptable level ($I^2=60%$).

When weighted for methodological quality, the effect size was less than 0.01-point difference (SMD=0.39, 95% CI [0.07-0.70]), suggesting that the overall effect was not significantly impacted by variation in methodological quality. No obvious asymmetry was identified from visual analysis of the funnel plot.

In sum, there were small effect sizes that favoured ACT over control conditions in improving psychological flexibility. However, problematic variation was observed and one study was highlighted as being overly influential. Thus, the effect size should be considered as falling within the range of 0.25-0.39.

4.7 Functioning

4.7.1 Self-reports

Six of the studies included a self-report measure of functioning (n=307). Four of these studies measured this outcome at follow-up (n=136). The average follow-up period was 2.75 months.

As seen in Figure 8, the random-effects model estimated a significant moderate effect (SMD=0.71, CI [0.30-1.11]). Heterogeneity was acceptable ($I^2=67\%$). The “one left out” analysis did not identify any study as exerting disproportionate influence on effect size. When studies were weighted for methodological quality, the meta-analytic effect increased by approximately 5% (SMD= 0.75, CI [0.33-1.17]). In terms of publication bias, the funnel plot did not indicate any concerns and the trim and fill procedure supported this.

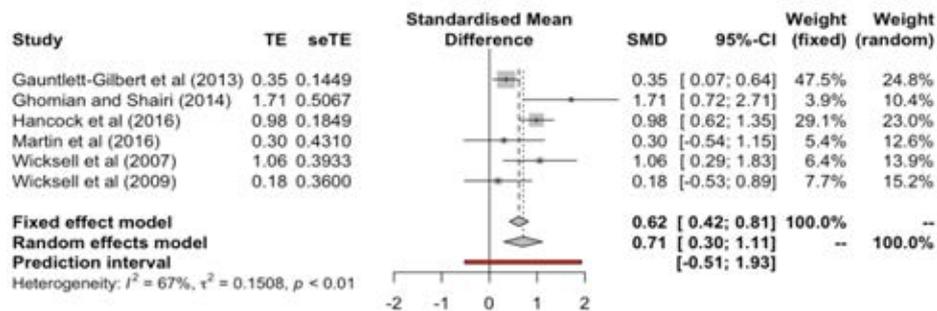


Figure 8. Forest Plot of Functioning Self-Report Post Intervention

In sum, a moderate effect was found to favour ACT improving functioning, as measured by self-report measures. This effect appeared robust to methodological variation and publication bias.

4.7.2 Follow-up

As seen in Figure 9, the random effects yielded a significant large effect in favour of the ACT improving functioning at follow-up (SMD=0.87, CI [0.24-1.49]). Heterogeneity falls just within acceptable levels ($I^2=71\%$).

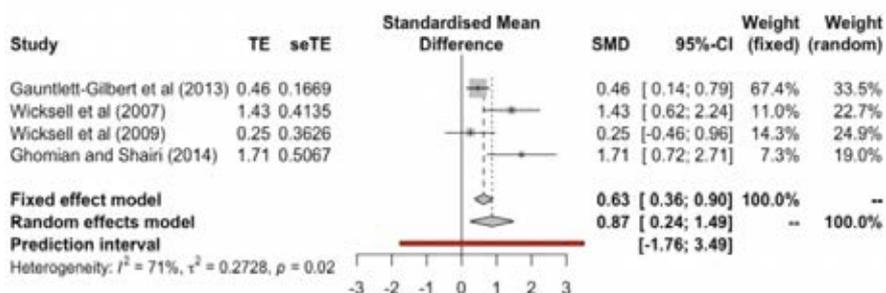


Figure 9. Forest Plot of Functioning Self-Report Follow-Up

The “one left out” analysis did not reveal any study as exerting disproportionate influence. When studies were weighted for methodological quality, there was a slight increase in estimated effect size (SMD=0.90, CI [0.26-1.54]). Visual inspection of the funnel plot is suggestive of publication bias. The trim and fill procedure estimated that one additional study needed to be added to the left side. The corrected model, as displayed in Figure 10, yielded a moderate effect size (SMD=0.66, CI [0.06-1.27]). Using the fail-safe N method, it is estimated that 36 studies would need to be published in order to make the result insignificant, indicating that the effect is robust to publication bias.

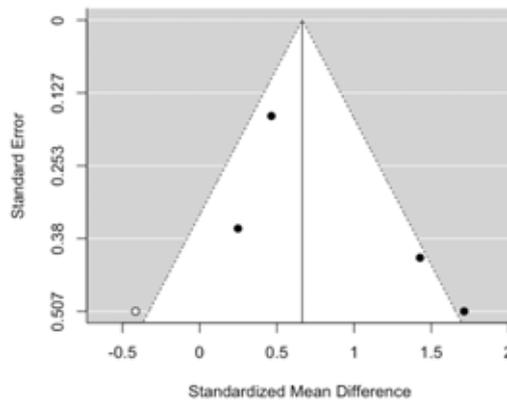


Figure 10. Funnel Plot of Functioning Self-Report Follow-Up

4.7.3 Reported by Parents

Functioning of children/adolescents as reported by parents was included in four studies (n=78), all of which used the Functional Disability Index (FDI) (Walker & Greene, 1991). All of these studies included participants with chronic pain and all of the interventions included parental involvement.

The random effects model produced a moderate but insignificant effect size in favour of ACT improving functioning, see Figure 11 (SMD=0.61, CI [-0.02-1.25]). A “one left out” analysis found that Martin et al (2016) was exerting disproportionate influence on meta-analytic effect. As a result, Martin et al (2016) was omitted and the results are shown in Figure 11

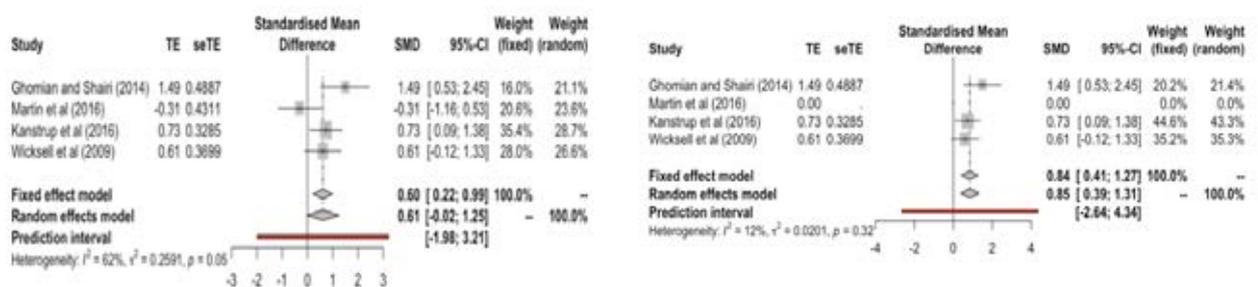


Figure 11. Forest Plots with and without Martin et al (2015)

Following omission of Martin et al (2016) the estimated meta-analytic effect was large and significant (SMD=0.85, CI [0.39-1.33]). Heterogeneity was deemed low ($I^2=12%$) and the “one left out” analysis did not reveal any one study disproportionately influencing effect size. When weighted for methodological quality, there was a small increase in effect (SMD=0.87, CI [0.40-1.33]). Visual inspection of the funnel plot did not indicate presence of publication bias.

In sum, the analysis estimated a large effect size in favour of ACT improving functioning of children/adolescents from the perspective of their parents.

4.8 Potential Moderating Variables

Analysis of potential moderating variables was undertaken to investigate the potential influence of study characteristics, clinical variables and sample demographics on meta-

analytic effect. The analyses highlighted a number of significant results, which varied across outcomes, see Table 7.

A consistent finding across all outcomes was the non-significant findings for study design, suggesting that this did not influence meta-analytic effect (all p s > 0.09). Another coherent finding that emerged across the literature was that Iranian studies reported disproportionately larger effect sizes in favour of ACT that were inconsistent with studies reported from other countries. As such, Iran was compared to the 'rest of the world' to further assess these differences. Iranian studies were found to have significantly larger effect sizes compared to the rest of the world in Anxiety (post-intervention) ($p < .001$) and Functioning (self-report post intervention) ($p=0.04$). An Iranian study was also omitted from Depression (post-intervention) for being disproportionately influential and two of the outcomes did not include Iranian studies. Thus, until the sources of variation within the Iranian studies has been identified, these studies should be interpreted with extreme caution as they are marked inconsistent with the literature from the 'rest of the world'.

The type of problem (physical health, mental health or prevention) targeted by the ACT was also found to elicit significant variation in effect sizes however, these effects were not consistent across outcomes. Within Depression (post-intervention), effect sizes were significantly larger for mental health (SMD=2.01, CI [0.64-3.38]) compared to prevention (SMD= 0.35, CI [0.16-0.87]) and physical health difficulties (SMD= 0.25, CI [0.04-0.55]). Whereas, for Anxiety (follow-up), larger improvements were observed for physical health (SMD= 0.53, CI [0.20-0.86]) compared to prevention (SMD= 0.16, CI [-0.03-0.35]). In Psychological Flexibility (post-intervention), larger improvements were seen for physical health (SMD= 1.44, CI [0.88-2.00]) compared to both mental health (SMD= 0.40, CI [0.07-

0.72]) and prevention (SMD= 0.07, CI [0.075-0.21]). Therefore, it may depend on what outcome is targeted as to which problem ACT is more effective for or alternatively these results may reflect a noisy data set. There did appear to be a trend of prevention programmes being less effective.

Parental involvement was shown to be a significant moderator that improved a number of outcomes including: Anxiety (follow-up) ($p=0.05$) and Psychological Flexibility (post-intervention) ($p=0.01$). Parental involvement was also observed to be approaching significance for Functioning Reported by Parents ($p=0.06$) and could not be calculated for the other functioning outcomes, due to all studies including parental involvement.

However, within the two outcomes where parental involvement was observed to be significant, the setting of the study was also a significant moderator. The studies that contributed to the larger effects included both parental involvement and were carried out in a community setting rather than a school setting. Consequently, it was not possible to detect which of these differences, setting or parental involvement was contributing to this change in effect.

The format of the intervention was found to be significant within the outcome category of Functioning Self-Report Follow-up. The individual format was shown to significantly increase effect size (SMD= 1.54, CI [0.91-2.17]) compared to a group format (SMD= 0.46, CI [0.14-0.79]). However, this was based on a small number of studies in each category (group, $n=1$; individual, $n=2$). This variable could not be calculated on four outcome categories, due to studies mainly offering ACT in a group format.

Age was not observed as a significant predictor of effect size. Whereas the percentage of females within a study was negatively associated with the effect size on two outcomes

(Functioning Self-Report Post Intervention, [$p=0.01$]; Functioning Self-Report Follow-up, [$p=0.00$]). Within Functioning Self-Report Follow-up, a significant negative correlation was also found between effect and year of study ($p=0.00$). However, this was not consistent across other outcomes and was based on a small number of studies. A significant negative correlation for length of intervention was also found for this outcome. However, it is possible this is an artefact of one study having a particularly large intervention length and showing non-significant effects (Gauntlett-Gilbert et al., 2013).

Table 7. Potential Moderator Variables

	Study Design (RCT/between-groups/within-groups)	Intervention Format (Group "G" Individual "I")	Parental Involvement (yes/no)	Country of study (Iran vs the rest of the world "RoW")	Problem Domain (Mental Health "MH" Physical Health "PH" preventative "PR")	Setting (community/school)	Age	Year of study	% Female	Length of Intervention
Depression Post Intervention (Extreme outlier studies removed)	P=0.73	P=0.06	P=0.18	N/A	P=0.39	P=0.17	P=0.66	P=0.97	P=0.98	P=0.66
Depression Follow-up	P=0.41	P=0.21	P=0.30	N/A	P=0.05 MH SMD= 2.01, [0.64-3.38] PH (SMD= 0.25, [-0.04-0.55]) PR (SMD= 0.35, [-0.16-0.87])	P=0.51	P=0.20	P=0.59	P=0.41	P=0.62
Anxiety Post Intervention	P=0.69	P=0.93	P=0.93	p<.001 Iran (SMD=2.77,[1.91-3.63]) RoW (SMD=0.26,[0.12-0.40])	P=0.88	P=0.93	P=0.09	P=0.38	P=0.98	P=0.04 Slope= -0.02

	Study Design (RCT/between-groups/within-groups)	Intervention Format (Group "G" Individual "I")	Parental Involvement (yes/no)	Country of study (Iran vs the rest of the world "RoW")	Problem Domain (Mental Health "MH" Physical Health "PH" preventative "PR")	Setting (community/school)	Age	Year of study	% Female	Length of Intervention
Anxiety Follow-up	P=0.10	N/A	P=0.05 Yes (SMD= 0.53, [0.20-0.86]) No (SMD= 0.16, [-0.03-0.35])	N/A	P=0.05 PH (SMD= 0.53, [0.20-0.86]) PR (SMD=0.16, [-0.03-0.35])	P=0.05 Community (SMD= 0.53, [0.20-0.86]) School (SMD= 0.16, [-0.03-0.35])	P=0.35	P=0.33	P=0.06	P=0.49
Psychological Flexibility Post Intervention	P=0.61	N/A	P=0.01 Yes (SMD= 0.81, CI [0.30-1.33]) No (SMD=0.10,[-0.09-0.28])	P=0.25	p<.0001 PH (SMD=1.44, [0.88-2.0]) MH (SMD=0.40, [0.07-0.72]) PR (SMD=0.07, [-0.08-0.21])	P=0.01 Community (SMD=0.81, [0.30-1.33]) School (SMD=0.10, [-0.09-0.28])	P=0.29	P=0.09	P=0.17	P=0.39
Functioning Self-Report Post Intervention	P=0.81	P=0.13	N/A	P=0.04 Iran (SMD=1.71, [0.72-2.71]) RoW (SMD=0.60, [0.21-0.97])	P=0.23	N/A	P=0.08	P=0.82	P=0.01 slope=-0.03	P=0.51
Functioning Self-Report Follow-up	P=0.09	P=0.006 I (SMD= 1.54, [0.91-	N/A	P=0.13	N/A	N/A	P=0.27	P=0.00 slope=-0.22	P=0.00 slope=-0.01	P=0.23

	Study Design (RCT/between- groups/within- groups)	Intervention Format (Group “G” Individual “I”)	Parental Involvement (yes/no)	Country of study (Iran vs the rest of the world “RoW”)	Problem Domain (Mental Health “MH” Physical Health “PH” preventative “PR”)	Setting (community/school)	Age	Year of study	% Female	Length of Intervention
		2.17) G (SMD= 0.46, CI [0.14-0.79])								
Functioning Report by Parents Post Intervention (Extreme outlier studies removed)	P=0.32	N/A	P=0.14	P=0.06	P=0.14	N/A	P=0.17	P=0.72	N/A	P=0.14

All significant results are highlighted in green.

5. Discussion

5.1 Summary of findings

The current review aims to investigate whether ACT improves various health outcomes for children and adolescents, as well as the ACT specific processes thought to mediate these outcomes. While a systematic review has already been conducted (Swain et al., 2015) this is the first meta-analysis of this literature. There was also a refinement of the inclusion criteria of the previous review to accommodate the particular circumstances of this literature and improve the quality of studies included. Potential moderating variables of effect were also investigated to provide information relevant to clinical decision-making and future research.

Twenty studies were included in the meta-analysis, examining the four most commonly reported outcomes (Depression, Anxiety, Psychological Flexibility and Functioning). Where data was available, these outcomes were analysed at least two months following the end of the intervention and the Functioning outcome was measured using parent report measures. Overall, significant effects were found across all outcomes favouring ACT over control conditions, although the size of effects varied. Small effects were found for Depression, Anxiety and Psychological flexibility and these effects were sustained at follow-up. However, a number of these summary effects were found to be at the boundaries of significance and became non-significant when weighted for methodological quality (Depression Post Intervention; Anxiety Post Intervention). These results suggest that conclusions are tentative until more methodologically rigorous research is conducted.

Moderate-to-large summary effects were reported for Functioning. Functioning relate to measures assessing illness or mental health-related activity limitations and was mainly

measured by the FDI, which may explain the consistency of results. These results were maintained when controlling for a number of factors and was maintained when measuring functioning of the child/adolescent by parent report.

There were mixed findings regarding how effect sizes differed between post intervention and follow-up periods. There were relatively large increases in effect in functioning (+0.16) and depression (+0.22), suggesting that ACT is not only effective, but may increase over time. However, there was a slight decrease in anxiety (-0.07). Thus, it is difficult to draw any firm conclusions especially considering that follow-up outcomes were based upon a much smaller proportion of studies.

Although it was anticipated that heterogeneity would be high as a result of the variation in the methodologies of the studies and differences in participant characteristics, only two of the outcomes demonstrated unacceptable heterogeneity (Anxiety Post Intervention; Psychological Flexibility Post Intervention). Publication bias also had little impact upon the estimation of summary effects, with two outcomes indicating the potential presence of publication bias (Anxiety Follow-Up; Functioning Self-Report Follow-up).

Therefore, these results cautiously indicate that ACT is an effective approach for children and adolescents, particularly in improving functioning. In comparison, the ACT adult literature has produced comparable small-to-moderate effect sizes ($d=0.66$; (Hayes et al., 2006); $g=0.57$; (A-Tjak et al., 2015); $g=0.42$, (Öst, 2014); $g=0.42$; (Powers et al., 2009). These studies reported potential publication bias and found that effect sizes were moderated by lower methodological quality studies producing larger effects. Meta-analyses for CBT with children have also shown slightly larger moderate effect sizes for depression ($d=0.66$; (Arnberg & Ost, 2014) and anxiety ($d=0.61$; (Ishikawa, Okajima, Matsuoka, & Sakano,

2007). However, these results were based on a small number of participants and studies were significantly heterogeneous.

5.2 Potential Moderating Variables

A number of clinical considerations have arisen from investigations into how moderating variables may impact overall meta-analytic effect. Prior to discussion, it is worth noting that the investigation into potential moderators was limited given that the ratio of studies to covariates was below the recommended 10 studies per covariate (Borenstein et al., 2009).

There were inconsistent findings when looking at different problem domains (physical health, mental health and prevention) suggesting that this may be dependent on the type of outcome measured. However, larger effect sizes were demonstrated in Anxiety (post-intervention) and Psychological Flexibility for physical health problems compared to mental health and prevention programmes. Also, many of the pain-focused studies included functioning outcomes, which were shown to have larger effects than other outcomes. These results may result from pain being the most widely researched area amongst this population (Swain et al., 2015) and/or because functioning taps into ACT relevant processes, whereby greater pain acceptance is associated with better emotional and physical outcomes (Pielech et al., 2017).

However, it is noted that many of the pain-focused studies included parental involvement, which was found to be another significant moderator across a number of outcomes. Thus, it is not clear whether parental involvement is the cause of higher effect size or whether the higher proportion of pain studies is causing the variable of parental involvement to be a significant factor. However, studies that were not focused on pain that

included parental involvement (Hancock et al., 2016; Timko et al., 2015) also demonstrated positive effects and support previous literature outlining this as an important factor within the chronic pain population (Palermo & Chambers, 2005; Palermo, Valrie, & Karlson, 2014). Therefore, it would be useful to compare results of interventions that include parental involvement against those that have not.

Parental involvement was also confounded by the setting of the study. Studies that took place in community settings as opposed to school settings were shown to have larger effect sizes. However, it was unclear if this was due to the fact that they all included parental involvement. One possible reason for these results is that a number of school-based ACT programmes were led by teachers who had limited ACT training (Livheim et al., 2015; Van Der Gucht et al., 2016). Research demonstrated that psychologist led programmes have better outcomes compared to teachers (Hofmann, Sawyer, Witt, & Oh, 2010; Stice & Shaw, 2004). Thus, it may be worth investigating ACT-based interventions in schools comparing interventions delivered by Psychologists and teachers.

Contrary to the assertion in other reviews (Halliburton & Cooper, 2015) that age may be a moderating factor, this review did not find reliable evidence to support this. However, this current meta-analysis still only included a limited age-range, with only four of the studies including children below the age of 10. Therefore, more studies would be needed to investigate the impact of age, especially those including younger children. Also, it was difficult to investigate the impact of different developmental adaptations made for children, as the descriptions were vague or it was unclear if adaptations were made at all.

As previously mentioned, Iran was reported disproportionately larger effect sizes compared to countries across the rest of the world. It is unclear why these differences have

emerged and researchers should interpret these findings with caution until these reasons are known. These differences may be due to the intervention being substantially different, the way the study was conducted or individuals from Iran responding differently to ACT from the rest of the world. Notably, five Iranian studies were rated as having lower methodological quality compared to other studies. These studies tended to report randomisation, but it was not clear how this was implemented. Treatment fidelity was generally not reported and it was unclear who delivered the interventions or what level of the training they had. Thus, researchers should be cautious about basing expectations on these findings.

5.3 Clinical Implications

Overall, it is indicated that ACT could be an effective treatment for children/adolescents, particularly in improving functioning. The evidence is less convincing within prevention programmes, indicating the setting or level of experience of those implementing the programme may be an issue. It also seems that parental involvement may be an important factor in the effectiveness of the intervention.

As previously reported, the evidence was limited in regards to ACT specific measures (Swain et al., 2015). The summary effect for psychological flexibility (as measured by versions of the AAQ) was similar to those found for anxiety and depression and less than functioning. This finding is surprising given that psychological flexibility is the primary goal of ACT, which raises questions about the validity and reliability of psychological flexibility as a measurable concept. While strong validity and reliability are reported for the AAQ-II (Bond et al., 2011) some have expressed concerns about the measure being more strongly related to negative affect rather than ACT specific processes (Wolgast, 2014). This could explain why the effect for psychological flexibility was found to be similar to mental health outcomes.

However, these results do not rule out the possibility that psychological flexibility is mediating the change of effect in clinical outcomes (Ciarrochi et al., 2010). One of the studies in this review (Livheim et al., 2015) demonstrated that psychological flexibility mediated a reduction in depression symptoms. However, one might have expected the magnitude of effect for psychological flexibility to be higher given the moderate-to-large effects found for functioning. It has been suggested that improvements in ACT process measures have been associated with improvements in daily functioning (McCracken & Gutiérrez-Martínez, 2011; Wicksell et al., 2013). Thus, it would be useful for more of the studies to include ACT specific measures in order to conduct mediational analyses to better understand these relationships.

Additionally, ACT targets six processes, with only one overarching process being evaluated within this analysis. A lack of validated measures means that these are not measured as frequently. Three other studies included alternative ACT specific processes including mindfulness (Mindful Attention Awareness Scale (MAAS)) (Livheim et al., 2015) and acceptance (Chronic Pain Acceptance Questionnaire (CPAQ)) (Gauntlett-Gilbert et al., 2013; Martin et al., 2016). The results for these processes were mixed, with mindfulness showing a marginally significant increase (Livheim et al., 2015), acceptance found to be associated with six out of eight outcomes in one study (Gauntlett-Gilbert et al., 2013) and non-significant findings for acceptance in another (Martin et al., 2016). Therefore, it currently unclear how these processes relate and whether these are the active ingredients that make the intervention successful.

5.4 Limitations of the Meta-Analysis

A limitation of this study includes the small number of studies included within each outcome. This means the power of statistical analyses is limited and the conclusions made are tentative. There are a range of different outcomes reported by studies, raising questions about what outcomes should be measured. When the same outcomes were measured, studies used different outcome measures, raising questions about which measures are the most reliable and valid.

Only two of the twenty studies included active control conditions (Hancock et al., 2016; Hayes et al., 2011) and as a result, it is difficult to attribute the summary effects specifically to the ACT intervention. It is possible that these effects were the results of non-specific therapeutic factors such as therapeutic alliance or increased contact time, as cited as a potential confounding factor in one study (Livheim et al., 2015). While Hancock et al (2016) included both ACT and CBT groups, these were not directly compared. The exclusion of an active control group has also been a criticism of the adult literature (Hacker, Stone, & MacBeth, 2016; Öst, 2014). It is important that future trials include active control conditions to determine whether ACT is effective compared to other treatments.

5.4 Convergence with Other Reviews

The current review appeared to demonstrate consistent findings with the four other reviews examining this area (Coyne et al., 2011; Halliburton & Cooper, 2015; Murrell & Scherbarth, 2006; Swain et al., 2015). These findings indicate that ACT could be a useful approach with children and adolescents and the statistical synthesis of the literature adds weight to this argument. The current review raised some additional questions about the conceptual clarity of psychological flexibility and how this interacts with clinical outcomes.

In agreement with the other reviews, a number of methodological issues were highlighted. The results showed that weighting for methodological quality did appear to impact the results, causing some of the effect sizes to become non-significant. The review supported arguments made within the physical health domain (Murrell & Scherbarth, 2006; Pielech et al., 2017) outlining the importance of parental involvement and indications that this may be useful across other domains, although these findings were limited due to the number of studies included in each comparison. Despite the queries about how age might affect outcome, no evidence was found to support this.

5.5 Recommendations for Future Research

It is recommended that further research with children/adolescents uses randomised trials, active control groups and larger sample sizes to examine the effectiveness of ACT. Confidence would also be strengthened by increasing the number of studies in each problem domain and increased clarity across studies about what constitutes a good outcome measure. Many of the studies were unclear about the level of experience of those delivering the interventions or combined ACT with other interventions, reducing their potential validity. Better control and measurement of treatment validity is required.

In order to provide clinical guidance on what ages ACT may be suitable, studies could include younger children to explore suggestions that children as young as seven could benefit from ACT (Coyne et al., 2011). Additionally, it would be useful for studies to include consistent ACT specific measures to investigate whether ACT is targeting what is intended.

5.6 Conclusion

The applicability of ACT for children and adolescents is further substantiated, supporting Swain et al (2015). The review produced further refinement demonstrating small

significant effects across a range of outcomes and showing moderate-to large effects within functioning outcomes. However, this effect may decrease when further studies are published as a result of including better quality studies. The review has also provided some indications as to factors that may be influencing the outcomes of the intervention, although the statistical power of these analyses was limited. This review provides a summary of the current literature; further research studies within different domains under more tightly-controlled conditions would strengthen and further elucidate these findings.

References

- Arnberg, A., & Ost, L.-G. (2014). CBT for children with depressive symptoms: a meta-analysis. *Cognitive Behaviour Therapy, 43*(4), 275–288.
- A-Tjak, J. G. L., Davis, M. L., Morina, N., Powers, M. B., Smits, J. A. J., & Emmelkamp, P. M. G. (2015). A meta-analysis of the efficacy of acceptance and commitment therapy for clinically relevant mental and physical health problems. *Psychotherapy and Psychosomatics, 84*(1), 30–36.
- 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.
- Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., ... Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy, 42*(4), 676–688.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons, Ltd.
- 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.

- Burckhardt, R., Manicavasagar, V., Batterham, P. J., Hadzi-Pavlovic, D., & Shand, F. (2017). Acceptance and commitment therapy universal prevention program for adolescents: a feasibility study. *Child and Adolescent Psychiatry and Mental Health, 11*, 27.
- Cartwright-Hatton, S., Roberts, C., Chitsabesan, P., Fothergill, C., & Harrington, R. (2004). Systematic review of the efficacy of cognitive behaviour therapies for childhood and adolescent anxiety disorders. *British Journal of Clinical Psychology, 43*(4), 421–436.
- Ciarrochi, J., Bilich, L., & Godsell, C. (2010). Psychological flexibility as a mechanism of change in acceptance and commitment therapy. *Assessing Mindfulness and Acceptance Processes in Clients: Illuminating the Theory and Practice of Change*, 51–75.
- Coyne, L. W., McHugh, L., & Martinez, E. R. (2011). Acceptance and Commitment Therapy (ACT): Advances and Applications with Children, Adolescents, and Families. *Child and Adolescent Psychiatric Clinics, 20*(2), 379–399.
- Doi, S. A. R., & Thalib, L. (2008). A Quality-Effects Model for Meta-Analysis. *Epidemiology, 19*(1), 94.
- Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health, 52*(6), 377–384.
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical Research Ed.), 315*(7109), 629–634.

- Gauntlett-Gilbert, J., Connell, H., Clinch, J., & McCracken, L. M. (2013). Acceptance and values-based treatment of adolescents with chronic pain: Outcomes and their relationship to acceptance. *Journal of Pediatric Psychology, 38*(1), 72–81.
- Ghomian, S., & Shairi, M. (2014). The effectiveness of acceptance and commitment therapy for children with chronic pain (chact) on the function of 7 to 12 year-old children. *International Journal of Pediatrics, 2*(3.1), 195–203.
- Greco, L. A., Blackledge, J. T., Coyne, L. W., & Ehrenreich, J. (2005). Integrating Acceptance and mindfulness into treatments for child and adolescent anxiety disorders. In *Acceptance and mindfulness-based approaches to anxiety* (pp. 301–322). Springer.
- Greco, L. A., & Hayes, S. C. (2008). *Acceptance & mindfulness treatments for children & adolescents: A practitioner's guide*. New Harbinger Publications.
- Green, H., McGinnity, A., Meltzer, H., Ford, T., & Goodman, R. (2005). *Mental Health of Children and Young People in Great Britain, 2004: (557702010-001)* [Data set]. American Psychological Association. Retrieved May 8, 2018,
- Hacker, T., Stone, P., & MacBeth, A. (2016). Acceptance and commitment therapy—Do we know enough? Cumulative and sequential meta-analyses of randomized controlled trials. *Journal of Affective Disorders, 190*, 551–565.
- Halliburton, A. E., & Cooper, L. D. (2015). Applications and adaptations of Acceptance and Commitment Therapy (ACT) for adolescents. *Journal of Contextual Behavioral Science, 4*(1), 1–11.
- 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 & Adolescent Psychology*, 1–16.
- 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.
- Hayes, S., 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.
- Hayes, S., Pistorello, J., & Levin, M. E. (2012). Acceptance and Commitment Therapy as a Unified Model of Behavior Change. *The Counseling Psychologist*, 40(7), 976–1002.
- Hayes, S., Villatte, M., Levin, M., & Hildebrandt, M. (2011). Open, aware, and active: Contextual approaches as an emerging trend in the behavioral and cognitive therapies. *Annual Review of Clinical Psychology*, 7, 141–168.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107-128.
- Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., ... Cochrane Statistical Methods Group. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343(oct18 2), d5928–d5928.
- Hofmann, S. G., Sawyer, A. T., Witt, A. A., & Oh, D. (2010). The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 78(2), 169.

- Ishikawa, S., Okajima, I., Matsuoka, H., & Sakano, Y. (2007). Cognitive Behavioural Therapy for Anxiety Disorders in Children and Adolescents: A Meta-Analysis. *Child and Adolescent Mental Health, 12*(4), 164–172.
- James, A., Soler, A., & Weatherall, R. (2005). Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane Database of Systematic Reviews, 4*.
- Kallapiran, K., Koo, S., Kirubakaran, R., & Hancock, K. (2015). Effectiveness of mindfulness in improving mental health symptoms of children and adolescents: a meta-analysis. *Child and Adolescent Mental Health, 20*(4), 182–194.
- Kanstrup, M., Wicksell, R. K., Kemani, M., Wiwe Lipsker, C., Lekander, M., & Holmström, L. (2016). A Clinical Pilot Study of Individual and Group Treatment for Adolescents with Chronic Pain and Their Parents: Effects of Acceptance and Commitment Therapy on Functioning. *Children, 3*(4), 30.
- Keenan, K., Feng, X., Hipwell, A., & Klostermann, S. (2009). Depression begets depression: Comparing the predictive utility of depression and anxiety symptoms to later depression. *Journal of Child Psychology and Psychiatry, 50*(9), 1167–1175.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 593–602.
- Kim, S. Y., Park, J. E., Lee, Y. J., Seo, H.-J., Sheen, S.-S., Hahn, S., ... Son, H.-J. (2013). Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *Journal of Clinical Epidemiology, 66*(4), 408–414.

- Levin, M. E., Hildebrandt, M. J., Lillis, J., & Hayes, S. C. (2012). The impact of treatment components suggested by the psychological flexibility model: A meta-analysis of laboratory-based component studies. *Behavior Therapy, 43*(4), 741–756.
- 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*(4), 1016–1030.
- Lundgren, T., Dahl, J., & Hayes, S. C. (2008). Evaluation of mediators of change in the treatment of epilepsy with acceptance and commitment therapy. *Journal of Behavioral Medicine, 31*(3), 225–235.
- Luoma, J. B., Hayes, S. C., & Walser, R. D. (2007). *Learning ACT: An acceptance & commitment therapy skills-training manual for therapists*. New Harbinger Publications.
- Martin, S., Wolters, P. L., Toledo-Tamula, M. A., Schmitt, S. N., Baldwin, A., Starosta, A., ... Widemann, B. (2016). Acceptance and commitment therapy in youth with neurofibromatosis type 1 (NF1) and chronic pain and their parents: A pilot study of feasibility and preliminary efficacy. *American Journal of Medical Genetics Part A, 170*(6), 1462–1470.
- McCracken, L. M., & Gutiérrez-Martínez, O. (2011). Processes of change in psychological flexibility in an interdisciplinary group-based treatment for chronic pain based on Acceptance and Commitment Therapy. *Behaviour Research and Therapy, 49*(4), 267–274.

- Moazzezi, M., Moghanloo, V. A., Moghanloo, R. A., & Pishvaei, M. (2015). Impact of Acceptance and Commitment Therapy on Perceived Stress and Special Health Self-Efficacy in Seven to Fifteen-Year-Old Children With Diabetes Mellitus. *Iranian Journal of Psychiatry and Behavioral Sciences*, 9(2).
- Moghanloo, V. A., Moghanloo, R. A., & Moazezi, M. (2015). Effectiveness of Acceptance and Commitment Therapy for Depression, Psychological Well-Being and Feeling of Guilt in 7 - 15 Years Old Diabetic Children. *Iranian Journal of Pediatrics*, 25(4).
- Murrell, A. R., & Scherbarth, A. J. (2006). State of the research & literature address: ACT with children, adolescents and parents. *International Journal of Behavioral Consultation and Therapy*, 2(4), 531.
- NHS England. (2016) *Implementing the five year forward view for mental health*. London: NHS England.
- Ollendick, T. H., Öst, L.-G., Reuterskiöld, L., & Costa, N. (2010). Comorbidity in youth with specific phobias: Impact of comorbidity on treatment outcome and the impact of treatment on comorbid disorders. *Behaviour Research and Therapy*, 48(9), 827–831.
- Öst, L.-G. (2014). The efficacy of Acceptance and Commitment Therapy: An updated systematic review and meta-analysis. *Behaviour Research and Therapy*, 61, 105–121.
- 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*, 18(8), 953–964.
- Palermo, T. M., & Chambers, C. T. (2005). Parent and family factors in pediatric chronic pain and disability: An integrative approach: *Pain*, 119(1–3), 1–4.

- Palermo, T. M., Valrie, C. R., & Karlson, C. W. (2014). Family and parent influences on pediatric chronic pain: a developmental perspective. *The American Psychologist*, *69*(2), 142–152.
- Pielech, M., Vowles, K. E., & Wicksell, R. (2017). Acceptance and Commitment Therapy for Pediatric Chronic Pain: Theory and Application. *Children*, *4*(2), 10.
- Powers, M. B., Vörding, M. B. Z. V. S., & Emmelkamp, P. M. (2009). Acceptance and commitment therapy: A meta-analytic review. *Psychotherapy and Psychosomatics*, *78*(2), 73–80.
- R Core Team. (2015). R: A language and environment for statistical computing. ISBN 3-900051-07-0.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.
- Rostami, M., Veisi, N., Dehkordi, F. J., & Alkasir, E. (2014). Benefits of Acceptance and Commitment Therapy, *2*(4), 8.
- Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D., & Patton, G. C. (2018). The age of adolescence. *The Lancet Child & Adolescent Health*, *2* (3), 223-228.
- Schwarzer, G. (2007). meta: An R Package for Meta-Analysis. *R News*, *7*, 40–45.
- Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-Analysis with R*. Springer International Publishing. Stice, E., & Shaw, H. (2004). Eating disorder prevention programs: a meta-analytic review. *Psychological Bulletin*, *130*(2), 206.

- 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.
- Terrin, N., Schmid, C. H., & Lau, J. (2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology*, *58*(9), 894–901.
- Theodore-Oklot, 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.
- Timko, C. A., Zucker, N. L., Herbert, J. D., Rodriguez, D., & Merwin, R. M. (2015). An open trial of Acceptance-based Separated Family Treatment (ASFT) for adolescents with anorexia nervosa. *Behaviour Research and Therapy*, *69*, 63–74.
- Van Der Gucht, K., W. Griffith, J., Hellemans, R., Bockstaele, M., Pascal-Claes, F., & Raes, F. (2016). Acceptance and Commitment Therapy (ACT) for Adolescents: Outcomes of a Large-Sample, School-Based, Cluster-Randomized Controlled Trial. *Mindfulness*, *8*, 1–9.
- Venta, A., Sharp, C., & Hart, J. (2012). The relation between anxiety disorder and experiential avoidance in inpatient adolescents. *Psychological Assessment*, *24*(1), 240.
- Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125.
- Walker, L. S., & Greene, J. W. (1991). The functional disability inventory: measuring a neglected dimension of child health status. *Journal of Pediatric Psychology*, *16*(1), 39–58.

- Wicksell, R. K., Kanstrup, M., Kemani, M. K., Holmström, L., & Olsson, G. L. (2015). Acceptance and commitment therapy for children and adolescents with physical health concerns. *Current Opinion in Psychology*, 2, 1–5.
- Wicksell, R. K., Melin, L., Lekander, M., & Olsson, G. L. (2009). Evaluating the effectiveness of exposure and acceptance strategies to improve functioning and quality of life in longstanding pediatric pain – A randomized controlled trial. *PAIN*, 141(3), 248–257.
- Wicksell, R. K., Melin, L., & Olsson, G. L. (2007). Exposure and acceptance in the rehabilitation of adolescents with idiopathic chronic pain – A pilot study. *European Journal of Pain*, 11(3), 267–274.
- Wolgast, M. (2014). What does the Acceptance and Action Questionnaire (AAQ-II) really measure? *Behavior Therapy*, 45(6), 831–839.

CHAPTER II: The Empirical Research Report

EXPLORING THE IMPLICIT RELATIONAL ASSESSMENT PROCEDURE (IRAP) AS A MEASURE OF COGNITIVE FUSION AND SELF-AS-CONTEXT IN STUDENTS

1. Abstract

1.1 Background

Cognitive Defusion and Self-As-Context are two processes in ACT that are thought to contribute to the overarching aim of psychological flexibility. These processes refer to a distancing from internal experiences and the ability to view these from an observer's perspective. There is research to suggest that these processes are effective in improving clinical outcomes, but the research does not demonstrate if they are working for the reasons stated and is overly reliant upon self-report measures.

1.2 Method

A student sample (N=77) was recruited to explore associations between self-reported cognitive fusion and psychological flexibility with psychological distress and self-esteem. The study piloted an implicit computer-based task called the Implicit Relational Assessment Procedure (IRAP) as a measure of fusion and self-as-context. Using Pearson's correlations associations between the IRAP and self-report measures were explored. A hierarchical regression was also conducted to investigate if the IRAP accounted for additional variance over the self-report measures.

1.3 Results

Positive associations were demonstrated between the ACT processes and psychological wellbeing measures. However, the response times on the IRAP did not correlate with any of the self-report measures and did not account for any additional variance.

1.4 Conclusion

Associations between the self-report measures indicated the processes targeted in ACT are justified. However, the predictive validity of the IRAP in assessing these processes was

poor. It was unclear whether this was due to confounding variables of the IRAP, the subclinical nature of the sample or because the ACT process terms need clarification.

2. Introduction

2.1 Overview of ACT and RFT

Acceptance and Commitment Therapy (ACT; Hayes, Strosahl & Wilson, 1999) is a mindfulness-based behavioural therapy, which aims to increase psychological flexibility. This involves accepting present-focused experiences of inevitable painful psychological processes, rather than attempting to suppress or avoid them in order to pursue value-driven behaviour (Foody, Barnes-Holmes, Barnes-Holmes, & Luciano, 2013; Hayes, Pistorello, & Levin, 2012). ACT attempts to achieve this through six interrelated processes: Cognitive Defusion, Acceptance, Being Present, Self-As-Context, Values and Committed Action (Harris, 2006). This study will focus specifically on the processes of cognitive defusion and self-as-context.

ACT is theoretically driven by Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001). RFT provides a behaviour-analytic account of the relationship between cognition and language, which is used to predict and explain various human behaviours (Duff, Larsson, & McHugh, 2016). RFT suggests that our shared complex capacity to arbitrarily derive relations between stimuli, without direct experience or training to do so is the basis of human suffering (Bach & Moran, 2008; Hughes & Barnes-Holmes, 2016). For example, if a child learns A is related to B in some way, then they will automatically be able to say that B is also related to A (see Hayes et al., 2001). RFT proposes this relationship between the way individuals relate to stimuli will alter their behaviour. Three ways of relating which shall be discussed further include; Co-ordination (A is the same as B) Hierarchical (A is part of B) and Deictic/Perspective-Taking (states the relationship from the speaker's perspective, "I/you", "here/there", "there/then" (Montoya-Rodriguez, Molina, & McHugh, 2016).

2.2 Cognitive Defusion and Self-As-Context Processes

This pattern of relating can cause suffering in the way we relate to our internal experiences, which implicates the development of self-concept. For instance, relationships of equivalence may be inferred between internal representations of experience and external reality. Thus, an individual who has the thought “I am worthless” would in this case place “I” and “worthless” in a relationship of equivalence, likely leading to increased distress and social withdrawal. A sense of self that is merged with the content of thoughts and feelings is described as cognitive fusion within ACT. This is thought to trap people into developing negative self-descriptions which are disconnected with ongoing external experience and hinders behaviour (Foody, Barnes-Holmes, & Barnes-Holmes, 2012).

ACT attempts to disrupt this co-ordination between self and verbal relations to facilitate a more defused perspective by helping individuals to view their internal experiences as harmless transient states (Hayes, Villatte, Levin, & Hildebrandt, 2011). There are a range of exercises and metaphors that are typically used to facilitate defusion including; using the statement “I am having the thought that” to highlight the non-literal quality of the thoughts (Healy et al., 2008).

Self-as-context is an interlinked process with cognitive defusion and is reliant upon the more basic process of defusion. However, Self-as-context is more explicitly related to the self and perspective-taking. This process aims to facilitate a ‘transcendent self’, referred to as the observing self. The self is viewed as a place that contains these internal experiences (“I am bigger than the thought”), rather than these forming their self-identity (Harris, 2006; Luoma, Hayes, & Walser, 2007) (See Figure 1).

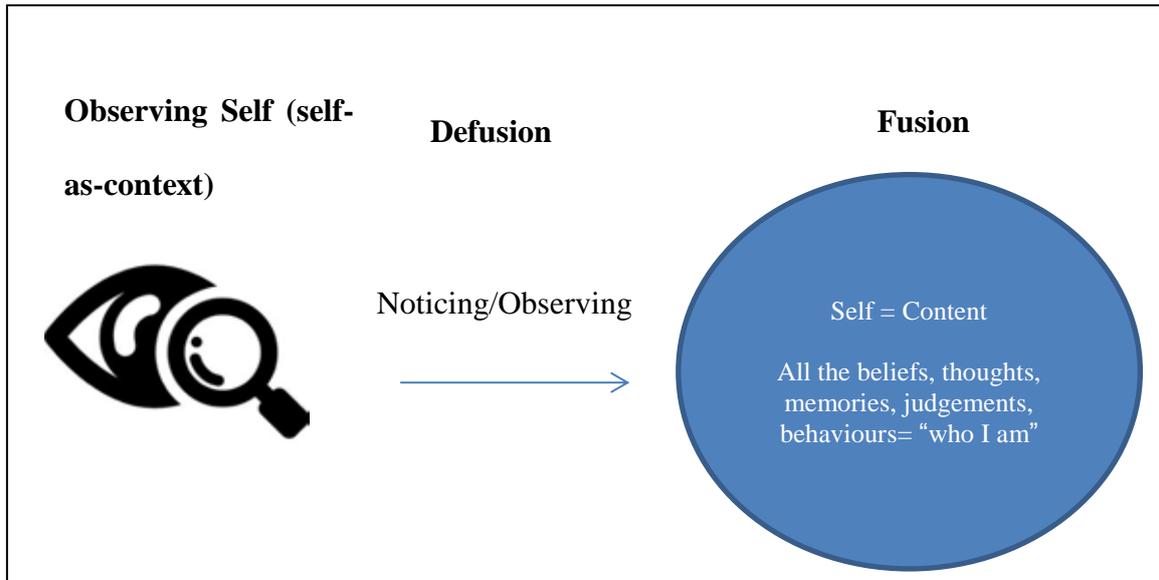


Figure 1. Defusion/Fusion and Self-As-Context Processes

From an RFT perspective, deictic (perspective-taking) and hierarchical relational responding are reported to underpin these processes (Luciano, Valdivia-Salas, & Ruiz, 2012; Luciano et al., 2011). It is proposed that individuals fused with thoughts relate as if both self and the content of internal experiences are in the **HERE** and **NOW**. Whereas, an individual with a more defused stance would view the self as being in the **HERE** and **NOW** and the content as being in the **THERE** and **THEN**. Hierarchical relational framing is thought to underpin the self-as-context process, whereby the self is viewed as higher in relation to their psychological content (Harris, 2006). Thus, both processes aim to create detachment from internal experiences but the self-as-context process explicitly highlights a hierarchical relationship.

2.3 Previous Research and Limitations

Research has shown positive outcomes for using techniques based on both these processes (Hayes, Luoma, Bond, Masuda, & Lillis, 2006; Levin, Hildebrandt, Lillis, & Hayes, 2012). However, there is less evidence to support whether they are effective for the reasons stated. Exploring ‘why’ an intervention is effective is becoming increasingly common within Clinical Psychology as opposed to just whether an intervention is effective (Barnes-Holmes, Hussey, McEnteggart, Barnes-Holmes, & Foody, 2016).

Consequently, there has been increasing focus on the associations between the six components targeted within ACT and levels of psychological wellbeing. For instance, Duff et al., 2016 found cognitive fusion and experiential avoidance was positively correlated with psychological distress, which was mediated by the believability of negative self-statements. Whereas Marshall et al., 2015 found that being more accepting of negative self-evaluations was a protective factor of mental wellbeing. Thus, these studies appear to support the use of defusion techniques aimed towards self-focused statements.

However, these studies have been criticised for their overreliance on self-report measures, highlighting a number of concerns around socially desirable responding (Power, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009), the limits of relying on self-reports as a substitute for observable behaviour (Kishita & Shimada, 2011) and relying on individuals’ ability to recognise their thoughts and how they relate to them (Vahey, Nicholson, & Barnes-Holmes, 2015).

There have also been concerns about how ACT processes are measured. For example, some researchers have argued that the Action and Acceptance Questionnaire-Version II (Bond et al., 2011) is more accurately a measure of negative affect than it is of psychological

flexibility (Wolgast, 2014). Whereas, there is a lack of adequate measures for the self-as-context process, which may be a reflection of this being something that people do not naturally report in regards to their experiences (Yu, Norton, Almarzooqi, & McCracken, 2017).

With self-as-context being difficult to measure, a number of studies have attempted to measure the impact of the hierarchical relational frames underlying this process. Experimental studies (Barnes-Holmes et al., 2016; Foody et al., 2013; Gil-Luciano, Ruiz, Valdivia-Salas, & Suárez-Falcón, 2017) and a single-case design (Ruiz & Perete, 2015) measured the impact of exercises using deictic relations to facilitate defusion compared with an exercise encouraging hierarchical relations between self and content. Findings demonstrated superior outcomes for exercises that included hierarchical framing between self and content. While this area of research is promising many of the studies utilised experimentally-induced distress as part of the protocol and there were difficulties in identifying what interactions were responsible for the observed changes.

2.4 The IRAP

One tool that has been developed based upon RFT principles is the Implicit Relational Assessment Procedure (IRAP) (Barnes-Holmes et al., 2006; Hughes & Barnes-Holmes, 2013). The IRAP was designed to capture ‘relational frames in flight’ (Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010) and the relative strength of these. The IRAP is a computer-based programme that requires opposing patterns of responding across successive blocks. The task requires participants to respond under time pressure and subtle differences in speed of responding across trials are measured. It is assumed that quicker responses will reflect relational frames that are more coherent with an individual’s learning history.

Thus, the IRAP provides a behavioural means of assessing cognitive processes that are potentially outside of the person's conscious awareness (De Houwer, 2006; Nosek, Hawkins, & Frazier, 2011) and assesses this through behavioural probability rather than through a mental construct (Hayes, Barnes-Holmes, & Wilson, 2012). The IRAP has been used increasingly in the ACT field, as a measure of deictic relational responding (Barbero-Rubio, López-López, Luciano, & Eisenbeck, 2016; Kavanagh, Barnes-Holmes, Barnes-Holmes, McEnteggart, & Finn, 2018), an assessment tool for a measure of change in defusion (Kishita, Muto, Ohtsuki, & Barnes-Holmes, 2014) and an assessment tool for experiential avoidance (Drake, Timko & Luoma, 2016). To the author's knowledge, the IRAP has not been used as a method of assessing hierarchical frames, which self-as-context processes are thought to be based upon.

2.5 The Current Study

The current study will explore the use of the IRAP as a measure of assessing cognitive fusion and self-as-context processes with a non-clinical student population. Currently, there is limited research investigating how individuals without psychological disorders relate to thoughts in terms of the RFT and ACT processes (Duff et al., 2016), which is considered important given that psychological difficulties are proposed to stem from our shared capacity for human language (Hayes et al., 2006). This may be helpful in identifying psychological processes that predispose individuals to psychological disorders, which could be used to inform ACT-based prevention programmes.

2.6 Aims and Hypothesis

Firstly, the aim of the study is to investigate whether the processes targeted in ACT are justified by investigating whether higher levels of fusion are associated with increased

psychological distress and lower self-esteem in the self-report measures. However, exclusive reliance on self-report measures is limited in what they can tell us about how individuals relate to their thoughts from an ACT perspective. Thus, secondly, a version of the IRAP will be piloted as an assessment tool of fusion and self-as-context. Participants will be required to respond to statements that represent either a co-ordinated relationship between self and content (representing fusion) or statements reflecting a relationship of containment between the self and content (self-as-context). Reaction times will be measured to detect biases across these conditions. Results from the IRAP will then be assessed for any correlations between any of the self-report measures. Thirdly, a hierarchical regression will be conducted to assess whether the IRAP can better predict psychological distress and self-esteem compared to the self-report measures. The following hypotheses are made:

Hypothesis 1: Individuals reporting higher levels of cognitive fusion and psychological inflexibility will report higher levels of psychological distress and lower self-esteem.

Hypothesis 2: Individuals who respond more quickly to statements reflecting more co-ordinated statements on the IRAP will report higher levels of cognitive fusion, psychological inflexibility, psychological distress and lower self-esteem. Whereas, faster reaction times on the containment trials are expected to correlate with lower scores on all the measures above apart from self-esteem, which would be expected to be higher.

Hypothesis 3: It is predicted that the IRAP may account for some additional variance compared with the self-report measures in predicting psychological distress and low self-esteem.

3. Method

3.1 Power Calculation

The sample size was based on a power calculation in another study using the IRAP with students and considering the high attrition rates of this procedure (Drake et al., 2016).

3.2 Participants

All procedures were approved by the University ethics board in March 2017 (See Appendix 1). Psychology students (N=77) were recruited from an UK university. The majority of the sample were female (n= 68, 89.4%), with a mean age of 20.51 years (SD=3.60). The largest ethnic groups represented were White/White British (n= 45, 58%) and Asian/Asian British (n=26, 34%).

All participants were recruited through the University's Research Participation Scheme whereby students are required to collect a certain number of participation credits as part of their course requirements. The study was advertised through an online system by an information sheet (see Appendix 2) and testing took place a minimum of 24 hours following the participant giving consent (see Appendix 3). The participants were subject to the inclusion/exclusion criteria in Table 1.

Table 1. Inclusion/Exclusion Criteria.

Inclusion Criteria	Rationale
The participants needed to be able to read and respond to questions/statements written in English.	All measures and questionnaires have been tested and validated in English and so, translating them may cause problems in terms of validity.
Aged 18 and over.	The target sample is an adult population, as there may be significant differences within younger participants. Also, participants are able to give informed consent at this age.
A student at the university.	To ensure consistency of testing settings and the practicalities of recruitment.
Exclusions Criteria	Rationale
Receiving support from mental health services, as identified from the screening questionnaire (i.e. under a Community Mental Health Team, Improving Access to Psychological Therapy Services or university counselling services).	The target sample is a non-clinical population who are not experiencing mental health difficulties.
Have a current mental health disorder diagnosis (i.e. schizophrenia, depression, anxiety, eating disorder etc.) or they currently experiencing mental health difficulties (i.e. depressive/anxiety symptoms), as identified by the screening questionnaire.	The target sample is a non-clinical population who are not experiencing mental health difficulties
Has a significant cognitive impairment/learning disability.	This may affect their capacity to carry out the questionnaires and IRAP. The speed of processing may impact on speed of respond on the IRAP and produce confounding effects.

3.2 Measures

3.2.1 The IRAP

The IRAP (Barnes-Holmes et al., 2010) is a computer-based task that is being piloted to assess whether an individual is relating in co-ordination (representing more fusion) or containment (representing a self-as-context stance) with positive and negative self-statements. The current study utilised a version of software for administering the IRAP programmed in 2009 (downloaded from <http://irapresearch.org/>).

The IRAP contains two stimulus elements, a target statement at the top of the screen referring to the nature to which they relate to something or not (See Table 2) followed by a target word appearing in the centre of the screen referring to hedonic tone (good versus bad, see Table 2). The statements are considered to represent co-ordinating self and a descriptor as

being equal (“I truly am”) or containing thoughts about being a certain way and recognising them as part of self (“I merely suspect I am”).

Table 2. Target Statements and Target Words Used in the IRAP.

Target Statements		Target Words	
Co-ordination	Containment	Positive	Negative
I really am	I only think I am	Good	Bad
I truly am	I just feel I am	Strong	Weak
I definitely am	I merely suspect I am	Nice Attractive Brave	Nasty Ugly Cowardly

This resulted in a 2x2 crossover design that produced four different trial-types including each combination of statement and word pairs. As seen in Figure 2, these trial-types include; co-ordination with positive self-statements (1), containment with negative self-statements (2), co-ordination with negative self-statements (3), containment with positive self-statements (4). The order in which labels and targets appeared was quasi-randomised, the only constraint being that the two types of target statement and two types of target words are presented together an equal number of times.

On each trial participants are required to confirm or deny the relationship between the target statements and target words by selecting one of two response options, “True” or “False” appearing at the right and left-hand corners of the screen by pressing the corresponding D and K keys. These response options were pre-determined by the instructions given (see Appendix 4).

The IRAP attempted to assess the strength of relationship with the different trial-type pairings by requiring opposing responses in successive blocks. For instance, in Block A, the

participant was required to respond “True” as if they were co-ordinated with positive descriptors (See 1 in Figure 2) and contained with negative descriptors (See 4 in Figure 2) and “false” to these statements in the subsequent Block B.

Thus, the variable of interest is the participant’s response times corresponding to these four trial-types, which is assumed will be inversely proportional to how personally relevant each of the trial-types are to the individual. All incorrect answers resulted in a red cross appearing in the centre of the screen and remained there until the correct response was selected. Correct responses cleared the screen and resulted in the next stimulus being presented after a 400ms delay.

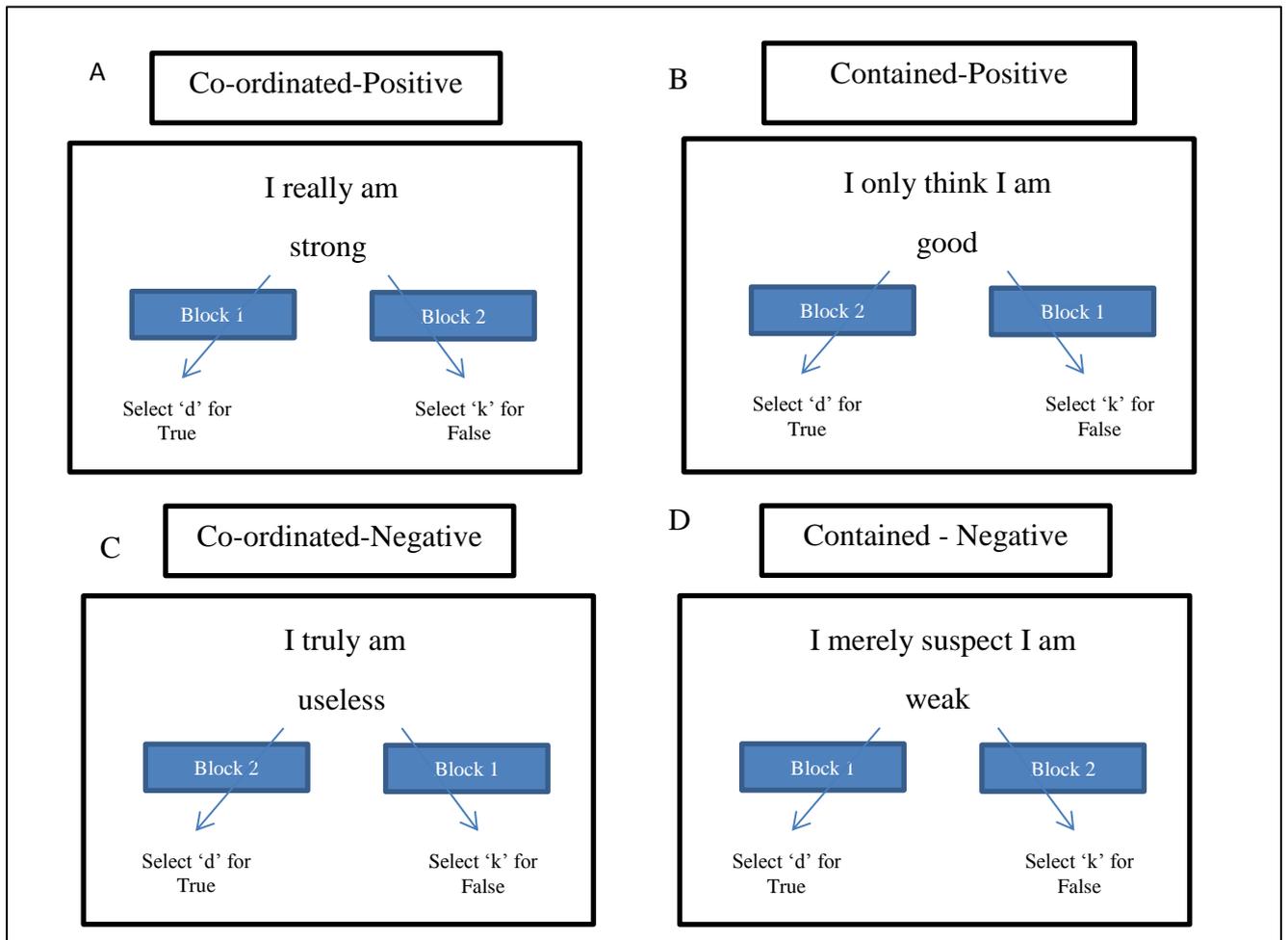


Figure 2. Trial Types and Required Responses for Block 1 and 2.

3.2.2 Psychological Distress

The Depression Anxiety Stress Scale (DASS 21) see Appendix 5 (Henry & Crawford, 2005) is a shortened version of Lovibond & Lovibond, 1995 42-item measure of Depression, Anxiety and Stress (DASS). Higher scores indicate higher levels of the three negative emotional states and each subscale is thought to tap into a more general dimension of psychological distress. Internal consistencies reported for the DASS-21 were .88 for Depression, .82 for Anxiety, .90 for Stress and .93 for the Total scale.

3.2.3 Self-Esteem

The Rosenberg Self-Esteem Scale (RSE) see Appendix 6 (Rosenberg, 1965) is a self-report measure evaluating global self-esteem. It includes 10 items measuring both positive and negative feelings about the self. Higher scores indicate higher levels of self-esteem. The scale has been reported to have an internal consistency of $a = .77$ (Rosenberg, 1965).

3.2.4 Cognitive Fusion

The Cognitive Fusion Questionnaire (CFQ) see Appendix 7 (Gillanders et al., 2014) is a seven item self-report measure of cognitive fusion. Higher scores indicate being more fused with cognition in general. There is a reported internal consistency at $a = .87$ (McCracken, DaSilva, Skillicorn, & Doherty, 2014).

3.2.5 Psychological Inflexibility

The Acceptance and Action Questionnaire, Version II (AAQ-II) see Appendix 8 (Bond et al., 2011) is a one-factor, seven-item measure of psychological inflexibility and experiential avoidance. Higher scores indicate higher levels of psychological inflexibility and experiential avoidance. The measure is reported to have an internal consistency of $a = .84$ and test-retest reliability of .81 (Bond et al., 2011).

3.2.6 Social Desirability Bias

The Marlowe-Crowne Social Desirability Scale (MCSDS) see Appendix 9 (Crowne & Marlowe, 1960) is a 33-item measure used to assess an individual's tendency to produce socially desirable responses. Lower scores indicate that participants are more willing to answer questions truthfully, even if they think those answers will meet disapproval. The scale is reported to have a test re-test reliability of 0.79 (Beretvas, Meyers, & Leite, 2002).

3.3 General Procedure

Experimental sessions took place in a small quiet room on the university campus conducted by one experimenter. All participants were required to read and sign a consent form at the beginning of the session, followed by completing a screening questionnaire (see Appendix 10). If participants were eligible to continue they would then fill out the demographic information sheet followed by all paper-based self-report measures. The order of the IRAP and self-report measures were not counter-balanced based on findings of a previous meta-analysis demonstrating that the order of implicit and explicit measures did not affect correlations (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005).

Following completion of the self-report measures the participant were given instructions about how to complete the IRAP procedure, which were adapted from a previous IRAP study (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008). Participants then proceeded to the practice blocks of the IRAP on a laptop computer. Participants were required to pass these two blocks (Block A and B) consecutively within the practice before being able to proceed to the testing block. In order to pass the block, participants were required to achieve an accuracy rate of 70% and an average response time of no more than 4000ms. Participants were allowed 3 attempts of the two pair of blocks. If they failed to meet these requirements after the three pairs of blocks then they would not complete any testing blocks.

Upon meeting the minimum requirements of the IRAP, participants proceeded onto the IRAP test blocks. Upon completion of the study, participants were debriefed about the study and credited for participating. Each data collection session took between 30 minutes to 1 hour, depending on how many practice blocks the participant was required to complete.

3.4 Data Preparation

Any test blocks that fell below 70% accuracy and an average response time of over 4000ms were excluded from the main analysis. If any participants scored below these requirements in more than one test blocks, then all of their IRAP data were removed from the analysis. This is based on similar procedures described in previous IRAP studies (Barnes-Holmes, Murphy, Barnes-Holmes, & Stewart, 2010; Barnes-Holmes, Waldron, Barnes-Holmes, & Stewart, 2009).

3.5 D Scores

The most commonly used method of quantifying differences between the IRAP trial-types was utilised. This method involves treating the differences in mean response latencies between Block A and Block B as an effect, using an adaptation of Cohen's standardised mean difference (Cohen, 1977). This measure is known as the *D-IRAP* score (for more details see, Barnes-Holmes et al., 2010; Greenwald, Nosek, & Banaji, 2003).

This calculation involves removing all outliers (latencies above 10,000ms) and calculates a difference score for each of the four trial-types. This is calculated by subtracting the mean response latency from all Block As to the corresponding Block Bs. Across all trials, if the *D-IRAP* score is greater than 0 then participants have responded more quickly during Block A compared to Block B. The trial-type scores were also inverted (multiplied by -1) in order to create a common axis across all trial-types, which makes them more interpretable (Hussey, Thompson, McEnteggart, Barnes-Holmes, & Barnes-Holmes, 2015).

3.6 Data Analysis

Data was analysed using SPSS Version 23 (IBM Corp, 2014). All descriptive statistics were compiled for each self-report measure, as well as the *D* scores for the IRAP. Scores of the completer of the IRAP and the non-completers were compared using *t*-tests. All self-report measures were subject to Pearson's correlation analysis to test whether they were associated (Hypothesis 1).

All *D-IRAP* scores for the four trial-types were compared using one sample *t*-tests. A repeated 4 x 2 ANOVA was conducted to see if there was a significant effect for trial-type.

A Pearson's correlation was conducted to examine the zero-order associations between all variables of interest (*D-IRAP* scores and all self-report measures; Hypothesis 2). A hierarchical regression was also conducted to test whether the IRAP trial-types accounted for additional variance above the self-report measures (Hypothesis 3).

4. Results

4.1 Attrition

Of 77 participants, 70 participants met the accuracy and speed requirements of the IRAP. Of the 7 participants, 5 of these failed to meet requirements in the practice block and so did not complete any test blocks. The remaining two completed the practice blocks, but failed at least two of the test blocks. The descriptive data for all of those who completed the IRAP and those who did not are provided in Table 3. One sample *t*-tests were conducted to assess differences in self-report measures between completers and non-completers of the IRAP, which were all found to be non-significant (all *ps* >0.34). A chi-square test was conducted for all other demographic variables comparing both the full sample and the sample that included only completers of the IRAP. No significant differences were found for any demographic variable (all *ps* >0.66).

The IRAP Non-completers average MCSDS score fell within the range of social desirability (>20), whereas the completers mean score was considered average. However, these differences were non-significant. Overall, 27% of the sample scored “high” in the MCSDS.

Table 3. Descriptive Statistics of Completers and Non-completers of the IRAP.

Measures	IRAP Completers (N=70)		IRAP Non-completers (N=7)	
	M ¹	SD ²	M	SD
CFQ	24.14	7.62	22.86	9.65
AAQ-II	19.9	7.33	16.86	7.60
RSE	19.59	3.94	18.71	4.50
DASS-21	12.46	7.46	11.86	8.86
DASS D	3.3	2.94	3.29	3.99
DASS A	3.06	2.80	2.71	3.30
DASS S	6.06	3.46	5.86	3.29
MCSDS	16.34	5.03	20.29	4.50
Age	20.51	3.53	21.71	4.03
Other Demographic Information				
	Full Sample (N=77)		Only Completers (N= 70)	
	N	%	N	%
Gender				
Males	8	11%	8	11%
Females	62	89%	55	79%
Education				
Postgraduates	13	17%	10	14%
Undergraduates	64	83%	60	86%
Ethnicity				
White/White British	45	58%	42	60%
Asian/Asian British				
Black/Black British	26	34%	23	33%
	7	9%	6	9%

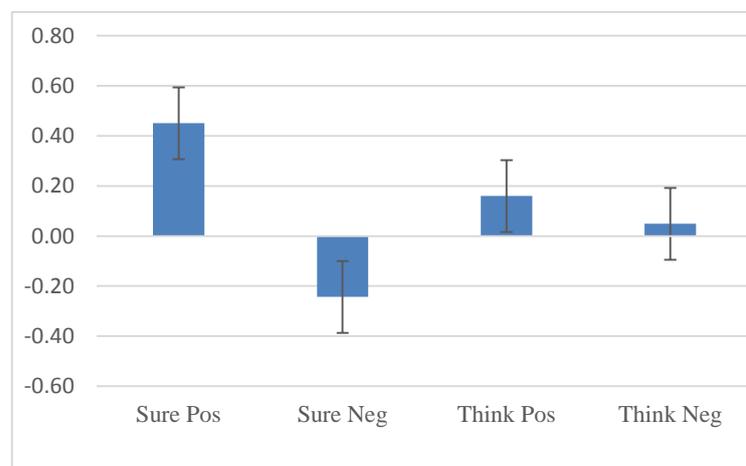
¹Mean ²Standard Deviation

4.2 IRAP Effects

All four IRAP trial-type scores were subject to one sample *t*-tests. As reported in the method, the IRAP latency data for 70 participants was transformed into four *D-IRAP* scores. A positive *D-IRAP* score represents quicker responding participants were asked to respond to

rule given in Block A (responding true to co-ordinated with positive words and Containment with negative words). Whereas, negative *D-IRAP* scores represent that respondents were quicker at responding to the rules given in Block B (responding true to co-ordinated with negative words and containment with positive words). Scores close to zero signify that there was no difference between responses given in Block A and Block B.

As seen in Figure 3, participants showed a significant bias towards responding “True” quicker than “False” when presented with co-ordinated statements with positive target words, $t(69) = 10.49, p < .0001$. Similarly, participants also gave a significantly quicker response when required to respond “True” when presented with co-ordinated statements with negative target words, $t(69) = -5.508, p < .0001$. Participants were much quicker at selecting “False” when presented with a containment statement with a positive target word, $t(69) = 3.48, p = 0.0009$. Participants showed a bias for responding “True” when presented with a containment statement with a negative word however this was not deemed significant, $t(69) = 1.06, p = 0.29$. Dependent t-tests comparing each of trial-types found that they were significantly different from one another ($p < .0001$), apart from the two containment trial types with each other, $t(69) = 1.65, p = 0.10$.

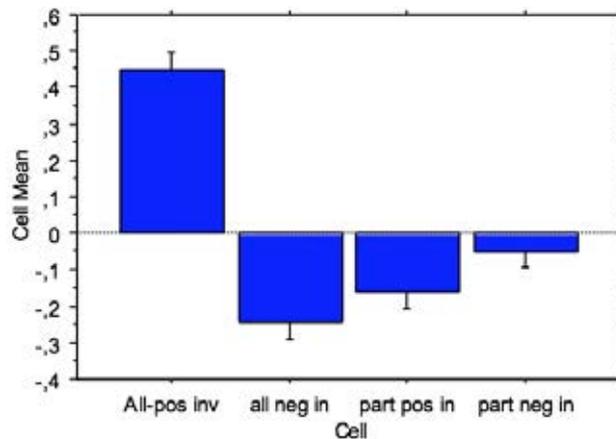


Sure = Co-ordination Trials; Think = Containment Trials

Figure 3. D-IRAP Scores

4.3 Inverted IRAP Scores

As mentioned, the last two trial-types were inverted for ease of interpretation. As such, positive *D-IRAP* scores reflect that a participant is endorsing a positive target word and negative *D-IRAP* scores reflect that the participant is endorsing negative target words in relation to themselves. As seen in Figure 4, the participants only appeared to endorse positive words about themselves in the co-ordinated positive trial-type. Whereas, they endorsed negative perception of themselves in all other trial-types. Thus, participants demonstrated that they were endorsing co-ordination with both positive and negative words, with a bias of perceived ‘negativity’ in all containment statements. For example, they do not contain positive thoughts, but they are able to contain some negative thoughts. A repeated 4 x 2 ANOVA showed that there was a main effect for inverted trial-types, $F(3,69) = 41.58, p < .0001$.



All = co-ordination trials; Part = containment trials

Figure 4. Inverted IRAP Trial-Types

4.4 Correlations between self-report measures (Hypothesis 1)

A number of significant correlations were found between related self-report measures, see Figure 5. This included negative associations between the RSE and CFQ ($p < .001$), AAQ-II ($p < .001$), DASS-21 ($p < .001$) and two of its subscales (DASS-21 D, $p < .001$ and DASS-21 A, $p < .05$). This showed that the more likely participants were to report their self-esteem as being higher the less likely they were to report higher fusion, psychological inflexibility and psychological distress. The CFQ demonstrated a positive association with the AAQ-II and DASS-21 including all of its subscales (all $ps < .01$). The MCSDS was negatively associated with the CFQ ($P < .001$), AAQ-II ($P < .05$) and the DASS-21 S ($P < .05$) with higher reports of social desirability associated with lower reported levels of fusion, psychological flexibility and DASS-21 S.

4.5 Correlations among self-report and *D-IRAP* scores (Hypothesis 2)

As seen in Figure 5, a significant association was observed between the co-ordinated positive trial-type and containment positive trial-type. Thus, the more likely the participant was to endorse that they were co-ordinated with positive words the less likely they were to reject that they contained positive words. There was also a positive association between co-ordination with positive target words and containment of negative target words. Therefore, the more likely participants were to endorse co-ordinated positive statements the more likely they were to agree that they contained negative descriptors. Only one significant correlation was found between *D-IRAP* scores and self-report measures, which was a negative association between the co-ordinated positive statements and the stress subscale of the DASS. The more likely that participants were to endorse co-ordination with positive target words, the less likely they were to report stress. No other significant correlations were found between trial-types and self-report measures ($Ps > 0.10$).

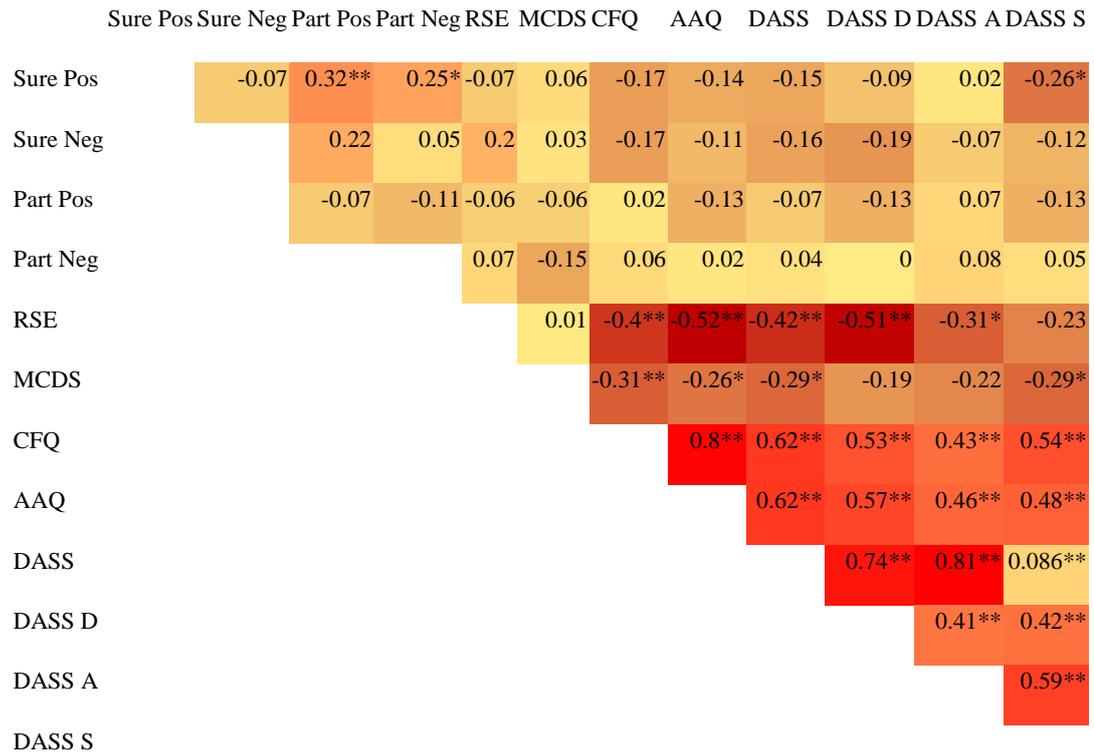


Figure 5. Correlations between IRAP Trial-Types and Self-Report Measures.

*Red denotes a stronger association, with lighter shades signalling weaker associations. Yellow denotes no/little difference in association. *= significant at the $p < .05$, **= significant at the $p < .01$. Sure = co-ordination trials, Part = containment trials*

4.6 Predictive validity of the IRAP (Hypothesis 3)

A hierarchical regression was conducted to assess the unique contribution of both the self-report measures and *D-IRAP* scores to psychological flexibility and cognitive fusion. This analysis was used to reduce the risk of a type one error resulting from collinearity. Two separate hierarchical regressions were conducted to investigate the predictive validity of measures for psychological distress (DASS-21) and self-esteem (RSE).

4.6.1 Psychological Distress

In predicting psychological distress, the CFQ and AAQ-II were entered at step one. As seen in Table 4, the CFQ and AAQ-II accounted for 43% of the variance ($R^2=.43$, $f(2,67) = 25.36$, $p=0.00$). The AAQ-II ($\beta = 0.34$, $p= 0.03$) and CFQ ($\beta = 0.34$, $p= 0.03$) were highlighted as significant predictors. When the four IRAP trial-types were entered at step two they only accounted for an additional 1% of the variance.

Table 4. Hierarchical Regression for Psychological Distress.

	<i>B</i>	<i>SE B</i>	β	R^2
<i>Step 1</i>				
CFQ	.314	.157	.32*	
AAQ-II	.35	.16	.34*	.43
<i>Step 2</i>				
CFQ	.31	.16	.32	
AAQ-II	.36	.16	.36*	
Co-ord Pos	-.51	2.24	-0.02	
Co-ord Neg	-0.87	2.02	-0.04	
Con Pos	-1.77	2.05	-0.09	
Con Neg	0.42	1.94	0.02	0.01 ^a

*denotes significance at $p < .05$ level. ^a R^2 Change

4.6.2 Self-Esteem

The CFQ and AAQ-II were entered into step one in the hierarchical regression for self-esteem. As seen in Table 5, these measures were shown to account for 27% of total variance ($R^2=0.27$, $F(2, 67) = 12.32$, $p= 0.00$). Psychological flexibility was found to be the only significant predictor ($\beta = -0.56$, $p= 0.002$). The IRAP trial-types were entered into step two and were only found to account for an additional 6% variance.

Table 5. Hierarchical Regression for Self-Esteem.

	<i>B</i>	SE <i>B</i>	β	R^2
<i>Step 1</i>				
CFQ	0.03	0.09	0.05	
AAQ-II	-0.30	0.09	-0.56*	0.25
<i>Step 2</i>				
CFQ	0.02	0.09	0.05	
AAQ-II	-0.30	0.09	-0.56*	
Co-ord Pos	-1.29	1.30	-0.12	
Co-ord Neg	1.66	1.18	0.16	
Con Pos	-1.09	1.19	-0.11	
Con Neg	0.94	1.13	0.09	0.06 ^a

*denotes significance at $p < .05$ level. ^a R^2 Change

5. Discussion

5.1 Summary of Findings

The current study adopted an inductive approach to investigate whether the underpinning processes targeted within ACT, specifically cognitive fusion and self-as-context are effective for the reasons stated. Thus, the study aimed to investigate whether individuals who were more fused and psychologically inflexible were more likely to be psychologically distressed and have lower self-esteem. The IRAP was also piloted as an alternative measure of these processes alongside self-report measures.

As predicted (Hypothesis 1) and in line with previous research (i.e. Duff et al., 2016; Marcks & Woods, 2005) strong associations were found between ACT process related self-report measures (CFQ, AAQ) and measures of psychological distress and self-esteem. The hierarchical regression found that AAQ-II was the only significant predictor of self-esteem.

However, as already reported a number of concerns have been raised about using self-report measures including the potential of social desirability bias (Powers, Vörding, & Emmelkamp, 2009). This appeared to be a valid concern, with 27% scoring within the ‘high range’ compared to the predicted 17% of participants expected to fall within this range (Crowne & Marlowe, 1960). Social desirability was also found to negatively correlate with the AAQ-II, CFQ and DASS-S indicating that individuals may have underscored these measures. Therefore, attempts to use an implicit measure appeared justified.

The IRAP results did not support hypothesis 2, that quicker responses on the coordinated trials would result in higher scores on self-report measures of cognitive fusion, psychological inflexibility, psychological distress and self-esteem. Similarly, the response times for the containment trials representing the self-as-context did not correlate with lower

scores on these measures. Unexpectedly, the only significant correlation demonstrated that those responding quicker on the co-ordination positive trial reported less stress. Thus, co-ordination with positive descriptors may not have adequately represented fusion and it is not possible to make a binary conclusion that where fusion is high containment will be low. The IRAP did not account for any additional variance than the self-report measures, disconfirming the final hypothesis.

In terms of responses to individual trial-types, individuals showed a significant bias towards responding quicker during the co-ordination trials, which appeared stronger for positive target words. Within containment trials, participants were more likely to disconfirm containment with positive words. However, participants responded quicker when agreeing with containment of negative words, although this was non-significant. Thus, all containment statements led to participants endorsing the negative views about themselves, whereas co-ordination facilitated both descriptors.

The IRAP data are difficult to interpret however, it is apparent that the predictive validity of the IRAP in this study to measure fusion and self-as-context processes was poor. A number of potential reasons that will be discussed are; limitations of the IRAP as measure of these processes, the delivery of the IRAP, the sample used and the difficulties in attempting to measure ACT's 'mid-level' terms.

5.2 Potential Confounding Variables of the IRAP Stimuli

Potential reasons for these unexpected results may be due to the stimuli used within the IRAP not accurately reflecting the ACT processes they intended to.

From an RFT perspective, these results could be explained in relation to theories of coherence whereby humans like things to make sense and coherence is assumed to be

reinforcing (Bordieri, Kellum, Wilson, & Whiteman, 2016). Co-ordination frames are the most basic and widely spread pattern of relational responding, which other relational frames are built upon (Blackledge, 2003; Hughes & Barnes-Holmes, 2016). Therefore, individual's learning histories are more likely to consist of co-ordination messages (i.e. You are good) rather than receiving messages containing hierarchical relations (i.e. I merely suspect that you are good). When there is an absence of contextual cues, individuals are more likely to relate to stimuli that involve lower levels of complexity (Barbero-Rubio et al., 2016; Hughes & Barnes-Holmes, 2016). Thus, faster responses to the co-ordination trial-types may have been the result of the sense-making of statements, rather than of how fused they were.

It is unclear whether co-ordinating more readily with descriptors reflected fusion, although the lack of association with the fusion questionnaire would suggest not. It would be interesting to examine the pattern of results using a clinical population to investigate whether this pattern was widespread or if it was affected by the presence of psychological difficulties.

While the IRAP was designed to measure relational frames, this was to the author's knowledge the first time that the IRAP has attempted to measure hierarchical relationships. Whereas previous studies have been used to assess what might be considered to be more 'simplistic' relational frames such as co-ordination/opposition frames (Barbero-Rubio et al., 2016). Hierarchical framing can involve increasingly complex interactions among relational frames and as a result the containment trials are much more demanding of participants and may take longer to process. Thus, it may not be suitable to compare these relational frames in this format.

Additionally, the stimuli within the IRAP may have been subject to various confounding variables impacting speed of response that were not equal across the four trials.

A number of potential variables have been highlighted in the IRAP research (Finn, Barnes-Holmes, Hussey, & Graddy, 2016; Finn, Barnes-Holmes, & McEnteggart, 2018) including the type of language used in stimuli across different trials, the coherence of statement and word combinations and biases of responding true or false.

For example, the statements used within the containment trials were much lengthier and more complex in nature, as well as using language that was probably less frequent in natural language (i.e. merely). The combination of statements and words would also be expected to be less common in natural language. Therefore, the participant's response times may have been confounded by the co-ordination trials containing more frequently used natural language opposed to responding more quickly because they were more fused with internal content. It would have been useful to conduct a bi-probe analysis in order to separate the differential effects of combining different statements and words together. However, this was not possible because the results yielded aggregate scores. With many confounding variables relating to language, it may be useful to include pictorial stimuli or more systematic testing of the language used within the stimuli to better control for this. For instance, it would be useful to assess the frequency of words used in the stimuli in the SUBTELX-NL lexical database, as used in word recognition tasks (e.g. Keuleers, Diependaele, & Brysbaert, 2010).

These findings may be used to explain why the co-ordinated positive trial produced such strong effects by arguably including more frequently used natural language. Thus, the fact that participants were able to show to occasional endorsement of negative self-statements may actually be a reflection of psychological flexibility. However, the containment negative trial-type did not correlate with higher self-esteem and lower psychological distress, as would be expected if this were true. Previous research has found that there is a bias in natural

language in responding true rather than false (Dodds et al., 2015), which may explain why three out of the four trials demonstrated a bias towards confirming statements.

5.3 Procedural Limitations

Other factors that may have influenced the results could include the way that the IRAP was implemented. The nature of instructions given has been highlighted as one such factor (Finn et al., 2016), although the exact implications of this are unclear (Finn et al., 2018). Findings demonstrated that providing general rather than specific rules appeared to elicit stronger IRAP effects and suggested that more detailed instructions may interfere with the automaticity of responding. Due to the complexity of the task and the high rates of attrition in previous studies (Drake et al., 2016), relatively detailed instructions were given during the practice blocks in this study. Although, the need to give such detailed instructions would pose questions about its clinical utility. Furthermore, the nature of instructions has also been found to interact with the order of trials (Finn et al., 2016). Therefore, the fact that block A was always presented first was another confounding variable and may have been reinforcing responses given in the co-ordination positive trial and containment negative trial, which was consistent with our findings. In one study that used limited instructions and counterbalancing across trials, the IRAP was found to be valid as a clinical assessment tool (Nicholson, McCourt, & Barnes-Holmes, 2013). Thus, future research may consider counterbalancing these trials and using more generalised instructions.

5.4 Nature of the Sample

While a non-clinical sample was selected to investigate processes that are thought to operate universally in humans, these participants would be expected to report lower levels on all self-report measures. Thus, the lack of associations found between self-report measures and between the IRAP may be a reflection of a lack of extreme variation in all domains. For

example, effect sizes for ACT were found to be much larger in outcome studies for individuals with more severe problems (Hayes et al., 2006).

The sample was mainly female and psychology university students, limiting the generalisability of these findings to gender and potentially higher than average Intelligence Quotient. There are also concerns that these participants are likely to have completed similar latency-based measures, which has been found to influence results in previous studies (Finn et al., 2016). Therefore, it would be useful to carry out the same procedure using a clinical population to investigate whether the results found are due to procedural/stimuli variables or the sample.

5.5 Difficulties in Measuring ACT Processes

The results of this study question the predictive validity of the IRAP in measuring the cognitive fusion and self-as-context processes. The findings showed that participants may be able to co-ordinate easily with self-related statements, but they may not necessarily be fused with them. Similarly, the containment trials may not be an adequate reflection of the person relating to themselves from a self-as-context perspective. While the six interrelated ACT processes have proved clinically useful they have been criticised as being middle-level terms. This refers to the terms being theoretically driven but non-technical and not generated within basic scientific research (Barnes-Holmes et al., 2016).

Consequently, these terms do not provide the precision that well-defined functional concepts do (i.e. reinforcement). For example, defusion may be used to refer a set of defusion procedures to produce an outcome, or it could be the process that these procedures are aiming to elicit or defusion may be viewed as an outcome. Thus, it is used in various ways and it is difficult to isolate what exactly defusion is. This may explain why the IRAP has been found

to be useful as a change measure of fusion following a defusion intervention (Kishita & Shimada, 2011), as this reinforces this circular definition. This may also explain why inconsistent and unexpected findings have been found when attempting to measure these different ACT processes (Drake et al., 2016). Thus, these concepts require further clarification, so that they can be adequately tested.

5.6 Conclusion

Overall, the findings of the self-report measures indicated that the processes targeted in ACT are justified, which was not replicated by the results of the IRAP. Further research is needed to examine whether this is because the IRAP is incompatible as a measure of these processes or whether the stimuli within the IRAP can be better adapted and controlled to measure these processes. The impact of conducting this study on a non-clinical population was unclear and it would be useful to compare these results obtained by a clinical population. However, prior to all of this, it has been suggested that the terms used within ACT need further redefinition so that they can be measured more precisely.

References

- Bach, P. A., & Moran, D. J. (2008). *ACT in Practice: Case Conceptualization in Acceptance and Commitment Therapy*. New Harbinger Publications.
- Barbero-Rubio, A., López-López, J. C., Luciano, C., & Eisenbeck, N. (2016). Perspective-taking measured by implicit relational assessment procedure (IRAP). *The Psychological Record, 66*(2), 243–252.
- Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist, 32*(7), 169–177.
- Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record, 60*(3), 527–542.
- Barnes-Holmes, D., Hayden, E., Barnes-Holmes, Y., & Stewart, I. (2008). The Implicit Relational Assessment Procedure (IRAP) as a response-time and event-related-potentials methodology for testing natural verbal relations: A preliminary study. *The Psychological Record, 58*(4), 497–516.
- Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2010). The Implicit Relational Assessment Procedure: Exploring the impact of private versus public contexts and the response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record, 60*(1), 57–79.
- Barnes-Holmes, D., Waldron, D., Barnes-Holmes, Y., & Stewart, I. (2009). Testing the validity of the Implicit Relational Assessment Procedure and the Implicit Association

- Test: Measuring attitudes toward Dublin and country life in Ireland. *The Psychological Record*, 59(3), 389–406.
- Barnes-Holmes, Y., Hussey, I., McEntegart, C., Barnes-Holmes, D., & Foody, M. (2016). The Relationship between Relational Frame Theory and Middle-Level Terms in Acceptance and Commitment Therapy. *The Wiley Handbook of Contextual Behavioral Science*, 365–382.
- Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement*, 62(4), 570–589.
- Blackledge, J. (2003). An introduction to relational frame theory: Basics and applications. *The Behavior Analyst Today*, 3.
- Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., ... Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, 42(4), 676–688.
- Bordieri, M. J., Kellum, K. K., Wilson, K. G., & Whiteman, K. C. (2016). Basic Properties of Coherence: Testing a Core Assumption of Relational Frame Theory. *The Psychological Record*, 66(1), 83–98.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences, Rev. ed.* Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349.
- De Houwer, J. (2006). What are implicit measures and why are we using them. *The Handbook of Implicit Cognition and Addiction*, 11–28.

- Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., ... Bagrow, J. P. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences*, *112*(8), 2389–2394.
- Drake, C. E., Timko, C. A., & Luoma, J. B. (2016). Exploring an implicit measure of acceptance and experiential avoidance of anxiety. *The Psychological Record*, *66*(3), 463–475.
- Duff, H., Larsson, A., & McHugh, L. (2016). Evaluations of self-referential thoughts and their association with components of acceptance and commitment therapy. *Journal of Contextual Behavioral Science*, *5*(1), 39–47.
- Finn, M., Barnes-Holmes, D., Hussey, I., & Graddy, J. (2016). Exploring the behavioral dynamics of the implicit relational assessment procedure: The impact of three types of introductory rules. *The Psychological Record*, *66*(2), 309–321.
- Finn, M., Barnes-Holmes, D., & McEntegart, C. (2018). Exploring the Single-Trial-Type-Dominance-Effect in the IRAP: Developing a Differential Arbitrarily Applicable Relational Responding Effects (DAARRE) Model. *The Psychological Record*, *68*(1), 11–25.
- Foody, M., Barnes-Holmes, Y., & Barnes-Holmes, D. (2012). The role of self in acceptance and commitment therapy. *The Self and Perspective Taking: Contributions and Applications from Modern Behavioral Science*, 125–142.
- Foody, Barnes-Holmes, Y., Barnes-Holmes, D., & Luciano, C. (2013). An Empirical Investigation of Hierarchical versus Distinction Relations in a Self-based ACT Exercise. *International Journal of Psychology and Psychological Therapy*, *13*(3).

- Gillanders, D. T., Bolderston, H., Bond, F. W., Dempster, M., Flaxman, P. E., Campbell, L., ... Remington, B. (2014). The Development and Initial Validation of the Cognitive Fusion Questionnaire. *Behavior Therapy, 45*(1), 83–101.
- Gil-Luciano, B., Ruiz, F. J., Valdivia-Salas, S., & Suárez-Falcón, J. C. (2017). Promoting Psychological Flexibility on Tolerance Tasks: Framing Behavior Through Deictic/Hierarchical Relations and Specifying Augmental Functions. *The Psychological Record, 67*(1), 1–9.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197.
- Harris, R. (2006). Embracing your demons: an overview of acceptance and commitment therapy. *Psychotherapy in Australia, 12*(4), 70.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. Springer Science & Business Media.
- Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual behavioral science: Creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioral Science, 1*(1–2), 1–16.
- Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). Acceptance and commitment therapy: Understanding and treating human suffering. *New York: Guilford*.
- Hayes, S., 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.

- Hayes, S., Pistorello, J., & Levin, M. E. (2012). Acceptance and Commitment Therapy as a Unified Model of Behavior Change. *The Counseling Psychologist, 40*(7), 976–1002.
- Hayes, S., Villatte, M., Levin, M., & Hildebrandt, M. (2011). Open, aware, and active: Contextual approaches as an emerging trend in the behavioral and cognitive therapies. *Annual Review of Clinical Psychology, 7*, 141–168.
- Healy, H.-A., Barnes-Holmes, Y., Barnes-Holmes, D., Keogh, C., Luciano, C., & Wilson, K. (2008). An Experimental Test of a Cognitive Defusion Exercise: Coping With Negative and Positive Self-Statements. *The Psychological Record, 58*(4), 623–640.
- Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44*(2), 227–239.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures. *Personality and Social Psychology Bulletin, 31*(10), 1369–1385.
- Hughes, S., & Barnes-Holmes, D. (2016). Relational frame theory. *The Wiley Handbook of Contextual Behavioral Science, 129–178*.
- Hughes, S. J., & Barnes-Holmes, D. (2013). A functional approach to the study of implicit cognition: The IRAP and the REC model. In *Advances in relational frame theory & contextual behavioural science: Research & applications* (pp. 97–126). CONTEXT PRESS.
- Hussey, I., Thompson, M., McEnteggart, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). Interpreting and inverting with less cursing: A guide to interpreting IRAP data. *Journal of Contextual Behavioral Science, 4*(3), 157–162.
- IBM Corp. (2014). IBM SPSS Statistics for Windows (Version 23). Armonk, NY: IBM Corp.

- Kavanagh, D., Barnes-Holmes, Y., Barnes-Holmes, P. M. D., McEnteggart, C., & Finn, M. (2018). Exploring differential trial-type effects and the impact of a read-aloud procedure on deictic relational responding on the IRAP. *The Psychological Record*.
- Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice Effects in Large-Scale Visual Word Recognition Studies: A Lexical Decision Study on 14,000 Dutch Mono- and Disyllabic Words and Nonwords. *Frontiers in Psychology, 1*.
- Kishita, N., Muto, T., Ohtsuki, T., & Barnes-Holmes, D. (2014). Measuring the effect of cognitive defusion using the Implicit Relational Assessment Procedure: An experimental analysis with a highly socially anxious sample. *Journal of Contextual Behavioral Science, 3*(1), 8–15.
- Kishita, N., & Shimada, H. (2011). Effects of acceptance-based coping on task performance and subjective stress. *Journal of Behavior Therapy and Experimental Psychiatry, 42*(1), 6–12.
- Levin, M. E., Hildebrandt, M. J., Lillis, J., & Hayes, S. C. (2012). The impact of treatment components suggested by the psychological flexibility model: A meta-analysis of laboratory-based component studies. *Behavior Therapy, 43*(4), 741–756.
- 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, 33*(3), 335–343.
- Luciano, C., Valdivia-Salas, S., & Ruiz, F. J. (2012). The self as the context for rule-governed behavior. *The Self and Perspective Taking: Research and Applications*, 143–160.
- Luciano, Ruiz, F. J., Vizcaíno Torres, R. M., Sánchez Martín, V., Gutiérrez Martínez, O., & López López, J. C. (2011). A Relational Frame Analysis of Defusion Interactions in

- Acceptance and Commitment Therapy. A Preliminary and Quasi-Experimental Study with At-Risk Adolescents. *International Journal of Psychology and Psychological Therapy*, 11(2).
- Luoma, J. B., Hayes, S. C., & Walser, R. D. (2007). *Learning ACT: An acceptance & commitment therapy skills-training manual for therapists*. New Harbinger Publications.
- Marcks, B. A., & Woods, D. W. (2005). A comparison of thought suppression to an acceptance-based technique in the management of personal intrusive thoughts: A controlled evaluation. *Behaviour Research and Therapy*, 43(4), 433–445.
- Marshall, S. L., Parker, P. D., Ciarrochi, J., Sahdra, B., Jackson, C. J., & Heaven, P. C. (2015). Self-compassion protects against the negative effects of low self-esteem: A longitudinal study in a large adolescent sample. *Personality and Individual Differences*, 74, 116–121.
- McCracken, L. M., DaSilva, P., Skillicorn, B., & Doherty, R. (2014). The Cognitive Fusion Questionnaire: A preliminary study of psychometric properties and prediction of functioning in chronic pain. *The Clinical Journal of Pain*, 30(10), 894–901.
- Montoya-Rodríguez, M. M., Molina, F. J., & McHugh, L. (2016). A Review of Relational Frame Theory Research Into Deictic Relational Responding. *The Psychological Record*, 67(4), 569–579.
- Nicholson, E., McCourt, A., & Barnes-Holmes, D. (2013). The Implicit Relational Assessment Procedure (IRAP) as a measure of obsessive beliefs in relation to disgust. *Journal of Contextual Behavioral Science*, 2(1–2), 23–30.
- Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences*, 15(4), 152–159.

- Power, P., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit Relational Assessment Procedure (IRAP) as a measure of implicit relative preferences: A first study. *The Psychological Record*, *59*(4), 621–640.
- Powers, M. B., Vörding, M. B. Z. V. S., & Emmelkamp, P. M. (2009). Acceptance and commitment therapy: A meta-analytic review. *Psychotherapy and Psychosomatics*, *78*(2), 73–80.
- Rosenberg, M. (1965). Rosenberg self-esteem scale (RSE). *Acceptance and Commitment Therapy. Measures Package*, *61*, 52.
- Ruiz, F. J., & Perete, L. (2015). Application of a relational frame theory account of psychological flexibility in young children. *Psicothema*, *27*(2).
- Vahey, N. A., Nicholson, E., & Barnes-Holmes, D. (2015). A meta-analysis of criterion effects for the Implicit Relational Assessment Procedure (IRAP) in the clinical domain. *Journal of Behavior Therapy and Experimental Psychiatry*, *48*, 59–65.
- Wolgast, M. (2014). What does the Acceptance and Action Questionnaire (AAQ-II) really measure? *Behavior Therapy*, *45*(6), 831–839.
- Yu, L., Norton, S., Almarzooqi, S., & McCracken, L. M. (2017). Preliminary investigation of self-as-context in people with fibromyalgia. *British Journal of Pain*, *11*(3), 134–143.

Chapter III: Public Dissemination Report

1. Introduction

Acceptance and Commitment Therapy (ACT) is a mindfulness-based behavioural therapy showing effectiveness across a range of conditions (i.e. depression, anxiety, chronic pain) (A-Tjak et al., 2015; Hayes et al., 2006). The primary goal of ACT is to live a meaningful life according to one's values, while accepting the presence of inevitable unwanted psychological processes (Harris, 2006).

ACT assumes that psychological distress is caused by shared human capacities and teaches general positive psychological skills (Lundgren, Dahl & Hayes, 2008). Given its proposed widespread applicability across difficulties and potential as a preventative intervention, interest in this approach has continued to grow.

The current thesis aims to contribute towards the evidence-base for ACT in two ways. Firstly, there will be a literature review of the published studies investigating the efficacy of ACT for children and adolescents. Secondly, an empirical research paper investigates two of the processes that are targeted within ACT to test whether they are justified and explores the use of an implicit measure in assessing these processes.

2. Literature Review:

The Effectiveness of Acceptance and Commitment Therapy with Children and Adolescents: A Meta-Analysis.

2.1 Background

A vast evidence-base demonstrates effectiveness of ACT with adults (A-Tjak et al., 2014; Hayes et al., 2006), but there is limited research with children and adolescents (Swain et al., 2015). This is surprising given the potential benefits of intervening at a young age and that the same psychological processes are thought to underpin their distress.

2.2 Method

A systematic search of online research databases was conducted to identify papers of studies investigating the effectiveness of ACT with children and adolescents. Twenty studies were identified and the four most commonly reported outcomes were included; 1) depression, 2) anxiety, 3) psychological inflexibility 4) functioning. A statistical procedure known as a meta-analysis was conducted to combine data from different studies to produce an overall score of effectiveness of ACT for each outcome. A quality criteria was specifically designed to assess areas of bias within studies and measured how this impacted the results.

2.3 Results

Results demonstrated that children and adolescents who received an ACT intervention reported a reduction in depression, anxiety and psychological inflexibility at the end of the intervention and at short-term follow-up. However, an analysis suggested that these results were not considered reliable when controlling for the quality of the studies. ACT demonstrated moderate-to-large effects in improving child and adolescent functioning, which remained stable when the quality of studies was measured and maintained at follow-up and

when measured by parent reports. Iranian studies reported that ACT demonstrated disproportionately larger improvements compared to other countries. It is unclear the reasons for this and so these studies should be interpreted with caution.

2.4 Conclusions

The review indicated that ACT was effective with children and adolescents. However, these findings were tentative given that some of these results were susceptible to change when controlling for study quality and included studies that showed disproportionate results. Thus, it would not be surprising if the estimate of overall effectiveness decreased when future studies were published. Better quality research studies are needed to strengthen these conclusions including comparing the ACT to other interventions, rather than control groups receiving no treatment.

3. Research Paper

Exploring the Implicit Relational Assessment Procedure (IRAP) as a Measure of Cognitive Fusion and Self-As-Context Processes in Students.

3.1 Background

Two therapeutic processes that are used within ACT to facilitate living a more meaningful life are cognitive defusion and self-as-context. Cognitive defusion aims to create a more detached stance from internal experiences (thoughts and feelings) by viewing them as harmless transient events. Self-as-context encourages a self-observing perspective of internal experiences and views themselves as the container of these experiences. The research supports that these techniques are effective, but it is unclear if this is for the reasons stated. Many of the studies rely on self-report measures, which relies on the individual's ability to access and accurately report relationships with their thoughts. There is a lack of adequate measures for assessing self-as-context process. Thus, the study seeks to explore the relationship between these ACT processes and psychological wellbeing. It will also explore the predictive validity of an implicit measure of these processes.

3.2 Method

A non-clinical student sample (N=77) was recruited from an UK University in order to assess associations between the ACT processes of cognitive fusion, measured by self-report compared with self-report measures of psychological distress and self-esteem. Participants also completed an implicit computer-based task aimed at measuring cognitive fusion and self-as-context processes. It attempted to do this by assessing how quickly participants responded true or false to statements appearing on a screen reflecting these processes. Response times were then statistically analysed to see if they were associated with scores on the self-report measures. A statistical analysis was conducted to test whether the implicit task was better at

predicting the psychological wellbeing outcomes compared to the ACT specific self-report measures.

3.3 Results

Individuals scoring higher on the cognitive fusion and psychological inflexibility self-report measures also reported higher levels of psychological distress and lower self-esteem. The implicit computer-based task did not appear to strongly correlate with any of the self-report measures and did not have any predictive power above the self-report measures.

3.4 Conclusions

Associations between the ACT self-report measures and psychological wellbeing outcomes suggested that the processes targeted in ACT are justified. However, these findings were not replicated in the implicit task. This suggests that the implicit task may not have been measuring what it was supposed to. Concerns were raised about confounding variables in the stimulus of the measure and the ability to measure processes that were not precisely defined.

4. Summary

The results demonstrate promise for the use of ACT with children and adolescents, although these are tentative and require further exploration. In understanding why ACT is showing effective results, the processes outlined in ACT do appear to be justified given the evidence from the self-report measures. However, an implicit measure failed to corroborate these findings and was not a valid measure of these processes. It was recommended that ACT processes need to be better operationalised before they can be measured accurately. There is still some way to go to understand how exactly the different components of ACT are contributing to overall effectiveness.

References

- A-Tjak, J. G. L., Davis, M. L., Morina, N., Powers, M. B., Smits, J. A. J., & Emmelkamp, P. M. G. (2015). A meta-analysis of the efficacy of acceptance and commitment therapy for clinically relevant mental and physical health problems. *Psychotherapy and Psychosomatics, 84*(1), 30–36.
- Harris, R. (2006). Embracing your demons: an overview of acceptance and commitment therapy. *Psychotherapy in Australia, 12*(4), 70.
- Hayes, S., 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.
- Lundgren, T., Dahl, J., & Hayes, S. C. (2008). Evaluation of mediators of change in the treatment of epilepsy with acceptance and commitment therapy. *Journal of Behavioral Medicine, 31*(3), 225–235.
- 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.

APPENDICES



Appendix 2. Participant Information Sheet



Title of Project: Exploring an Implicit Measure of Acceptance and Commitment Therapy (ACT) Processes of Cognitive Fusion and Self-As-Context in Students.

Researchers: Lois Carter (Chief Investigator, Trainee Clinical Psychologist), ***** (Supervisor, Clinical Psychologist)

We would like to invite you to participate in a research study. Below you will find some information about the study. Please read through the information carefully and if you have any queries/questions, please do not hesitate to contact us. The contact details are provided at the bottom of the last page.

What is the purpose of this research?

The purpose of the research is to explore whether the way we relate to our thoughts may impact our mood and how we feel about ourselves. For example, individuals with psychological difficulties, such as anxiety and/or depression can find it very difficult to separate themselves from the negative thoughts about themselves, believing them to be the literal truth. Acceptance and Commitment Therapy (ACT) aims to change an individual's relationship with their thoughts, so that they are less influenced by them. Thus, the study is looking to see if how we

associate with our thoughts relates to the levels of psychological distress and self-esteem that we experience, testing if ACT is effective for the reasons it is claimed to be.

Inclusion Criteria	Exclusion Criteria
<ul style="list-style-type: none">• Able to read and respond to questions written in English.	<ul style="list-style-type: none">• If you are receiving support from mental health services (i.e. under a Community Mental Health Team, Healthy Minds, counselling support etc.)

<ul style="list-style-type: none"> • Aged 18 and over. 	<ul style="list-style-type: none"> • If you have a current diagnosis of a mental health disorder (i.e. schizophrenia, bipolar disorder, post-traumatic stress disorder, obsessive-compulsive disorder, eating disorder, personality disorder etc.)
<ul style="list-style-type: none"> • A student at the University of Birmingham. 	<ul style="list-style-type: none"> • If you are currently experiencing mental health difficulties (i.e. depression, anxiety etc.) which will be identified via
	<ul style="list-style-type: none"> • If you have a significant cognitive impairment and/or learning disability.

Why have I been invited to take part?

You have been invited to take part in this study because you are a student at the University of Birmingham. This study is specifically interested in exploring the thinking styles of a non-clinical population, as ACT is proposed to be applicable to all. The full inclusion/exclusion criteria are listed in the table below.

What will happen to me if I agree to take part?

If you agree to take part in the study, you will be asked to sign a consent form. The researcher will arrange a time and date that is convenient for you and you will be asked to complete five short pen and paper questionnaires, followed by a computer task. Total participation will take around 1 hour.

What will happen if I do not want to carry on with the study?

Your participation is entirely voluntary. You can withdraw at any time during the research without having to give a reason. If you chose to withdraw, this will have no effect on your current or future relationship with the University of Birmingham. You can do this by contacting the researcher, contact details are provided at the bottom of the last page. Please be aware that the latest date that you can

withdraw your data will be January 2018. This deadline is in place so that the researcher has time to analyse and write-up the results, removal of data at this point would impact the overall analysis.

Expenses and payments

If you are enrolled in the University of Birmingham's Research Participation Scheme you will receive 1.0 academic credit. No other expenses or payments will be made.

What will happen to the results of the research study?

The results of the study will be written up for my Clinical Psychology Doctorate thesis. The findings may be submitted to a peer-reviewed journal for publication. You will not be identified in any report or publication. If you wish, we will send you a summary of the findings, once the study has been completed.

What are the possible disadvantages and risks of taking part?

You will be asked to reflect on your current levels of distress and answer questions relating to beliefs about yourself. While it is considered unlikely, there is a small risk that this could highlight difficult feelings that you may be struggling with, which may cause some anxiety or upset. If you feel uncomfortable at any time we encourage you to talk to someone close to you or to raise your concerns with the researcher. You may wish to seek additional support from someone at the University such as your Personal Tutor, University counselling service or GP.

There are also a number of helplines you can call confidentially should you need; these are provided below:

Samaritans

Website: www.samaritans.org

Tel: 08457 90 90 90

Get Connected

Website: www.getconnected.org.uk

Tel: 0808 808 4994

SANE

Website: www.sane.org.uk

Tel: 0300 304 7000

What happens if I have any further concerns?

If you have any concerns about how the study was conducted or questions that you were asked please contact ***** (Research Supervisor)

If you would like to discuss this research in any more detail, please contact me using the contact details below:

Lois Carter (Trainee Clinical Psychologist)

Email: *****

***** (Research Supervisor)

Telephone: *****

Email: *****

Appendix 3. Informed Consent Form



UNIVERSITY OF BIRMINGHAM

Participant Identification Number:.....

Title of Project: Exploring Acceptance and Commitment (ACT) processes of cognitive fusion and self-as-context in students.

Researcher: Lois Carter, Trainee Clinical Psychologist

Please initial box

I confirm that I have understood the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

1. I understand that my participation is voluntary and that I am free to withdraw at any time during the research, up until January 2018 without giving any reason, without effecting my current or future relationship with the University of Birmingham.

2. I understand that the completion of questionnaires and the computer-based task should take approximately 1 hour.

3. I understand that the data collected during this study will be looked at by the researcher and relevant others at the University of Birmingham to ensure that the analysis is a fair and reasonable representation of the data. Parts of the data may also be available to research collaborators.

4. I agree to take part in the above study.

.....

.....

.....

Name of participant

Date

Signature

.....

.....

.....

Name of researcher

Date

Signature

If you would like to receive information of the findings and any articles that are published from the research please mark the box

Appendix 4. IRAP Instructions

INSTRUCTIONS

The second part of the test is the computer task. First of all, its just about familiarising yourself with the task. If you could start by hovering your fingers over the “D” and “K” keys, this are the two main keys you will be using. If you press “D” this means that you are responding “True” and if you press the “K” key this means you are responding “False”.

Statements are going to appear on the screen which you will have to respond true or false to. You will not be responding in terms of how you feel about yourself but in relation to a rule that I give you. You can treat it a bit like a game, in trying to work out what the correct responses are to this rule and to try and get as many correct as you can. Take your time to work out the rule and the correct responses, this might not make sense initially, but it might start to make sense over time. Your goal is to avoid the red X’s.

To do this you need to take your time initially – if you take your time at the start of the practice phase the tasks become easy quickly (even if a little weird), but if you rush yourself at the start it becomes panicky and difficult – **so give yourself a chance and take it slowly, especially at the start.**

During the experiment, you will be asked to respond as **accurately** as you can across all trials. When you make an incorrect response for a task it is signalled by the appearance of a red ‘X’ in the centre of the screen. This will signal to you that you need to change your answer. If you see lots of red crosses appear, this may be a cue to slow down. Don’t worry if you make mistakes, just notice when you get the red cross and change the way you respond to this.

The first rule I would like to give you to respond to the statement as if you have a very high self-esteem, think very highly of yourself and you think things like I am really good/strong. However, you might have the occasional negative thought for example if something does not go your way, then you might have a fleeting thought that you are bad. However, you think of yourself mainly in a positive light.

So according to the rule that you have a very positive self-esteem with some partially negative thoughts, what would you respond to this trial with? and this one?

(Pointing to each of the four sample trials below to check that they understand how to respond in line with the rule. Also, if you were going to have the response options altering position during the task you would let people know that as you show them the sample trials.)

For the second block, I am now going to switch the rule. Now I want you to respond as if your self-esteem is really at rock-bottom, you believe you are worthless and have a negative view of yourself. However, you might get a fleeting positive thought about yourself, for example, if you do something well you might think maybe sometimes I might be okay but you think mainly negatively about yourself.

So according to the rule that you have a very positive self-esteem with some partially negative thoughts, what would you respond to this trial with? and this one?

(Pointing to each of the four sample trials below to check that they understand how to respond in line with the rule. Also, if you were going to have the response options altering position during the task you would let people know that as you show them the sample trials.)

When they reach the accuracy screen I have a look at their scores with them.

If they have reached an accuracy of >70% and latency of <4secs:

Great, so this task wants you to get an accuracy of 70% or above and to respond within 4000ms so your responding is perfect. Well done!

If they get good latency but poor accuracy:

Ok so your responding time is great because this task wants you to respond within 4000ms but it wants you to get an accuracy of 70% or above. So you can slow down as you have plenty of time there to play around with to get your accuracy up. Don't worry that's what the practice is for, to give you the time to get used to this.

If they get good accuracy but poor latency:

Ok so your accuracy is great because this task wants you to get an accuracy of 70% or above but it wants you to respond within 4000ms so you just need to speed up a little. Don't worry that's what the practice is for, to give you the time to get used to this and as you become familiar you'll get quicker.

Illustrations

Shown below are illustrations of the four different types of statement that will be presented repeatedly in this part of the experiment. To help you understand the meaning of each response is explained in the four illustrations below. Please examine each illustration and then read carefully the explanation attached to it – it's important to understand what you're response options mean before your task can be explained to you.

Illustration 1

I truly am

Strong

Select 'd' for True	Select 'k' for False
-------------------------------	--------------------------------

Explanation for Illustration 1

If you select "True" by pressing the 'D' key, you are agreeing with the statement "**I am strong**".
If you select "False" by pressing the 'K' key, you are disagreeing with the statement "**I am strong**".

Illustration 2

I only think I am

Useful

Select 'd' for True	Select 'k' for False
-------------------------------	--------------------------------

Explanation for Illustration 2

If you select "True" by pressing the 'D' key, you are agreeing with the statement "**I only think I am Useful**".
If you select "False" by pressing the 'K' key, you are disagreeing with the statement "**I only think I am Useful**".

Illustration 3

I am completely

Bad

Select 'd' for
True

Select 'k' for
False

Explanation for Illustration 3

If you select "True" by pressing the 'D' key, you are agreeing with the statement "I am completely bad."

If you select "False" by pressing the 'K' key, you are disagreeing with the statement "I am completely bad."

Illustration 4

I merely suspect that I am

Weak

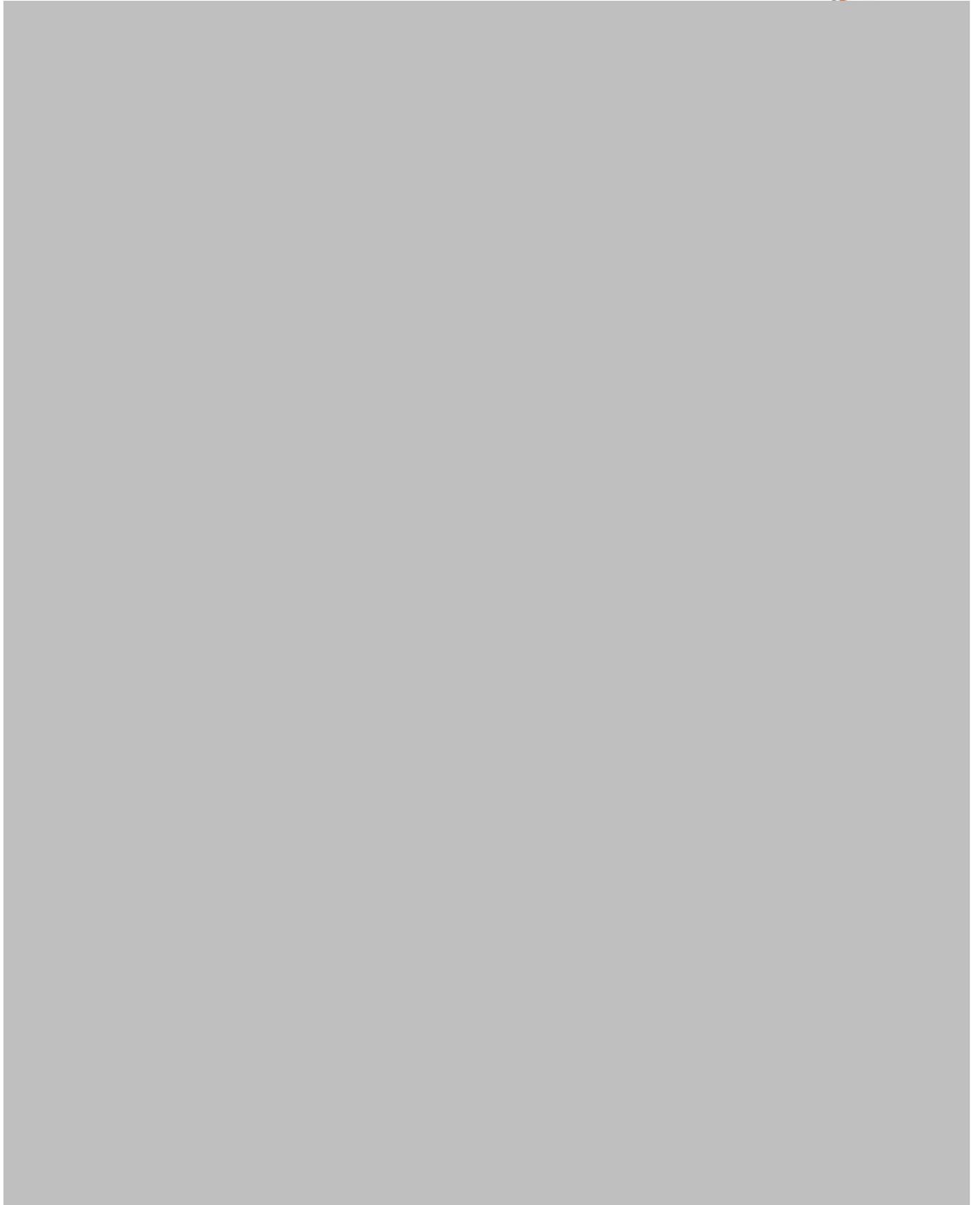
Select 'd' for
True

Select 'k' for
False

Explanation for Illustration 4

If you select "True" by pressing the 'D' key, you are agreeing with the statement "I merely suspect that I am weak"

If you select "False" by pressing the 'K' key, you are disagreeing with the statement "I merely suspect that I am weak"



Appendix 6. Rosenberg Self-Esteem Scale



Appendix 7. Cognitive Fusion Questionnaire



Appendix 8. Acceptance and Action Questionnaire- Version II



Appendix 9. Marlowe Crowne Social Desirability Scale



Appendix 10. Screening Questions



Screening Questions

The study is specifically seeking a non-clinical population. Thus, if you are currently experiencing mental health difficulties and or seeking support from mental health services then unfortunately you will not be eligible to participate in this study.

If you are currently experiencing mental health difficulties and feel you need support, there some contact numbers that you can call confidentially at the bottom of the page.

- 1) Are you currently receiving support from any mental health services? (i.e. Community Mental Health Teams, Healthy Minds, counselling support etc.)

Yes No Prefer not to say

- 2) Do you have a current diagnosis of a mental health disorder? (i.e. schizophrenia, bipolar disorder, post-traumatic stress disorder, obsessive-compulsive disorder, eating disorder, personality disorder etc.)

Yes No Prefer not to say

- 3) Are you currently experiencing any mental health difficulties? (i.e. depression, anxiety, psychosis etc.)

Yes No Prefer not to say

Please tick this box to indicate that you are not experiencing any of the above issues

Only if you have ticked this box and have answered no to all of the above, then you will be eligible to continue with participation in the study.

Samaritans

Website: www.samaritans.org

Tel: 08457 90 90 90

Get Connected

Website: www.getconnected.org.uk

Tel: 0808 808 4994

SANE

Website: www.sane.org.uk

Tel: 0300 304 7000