VOLUME ONE

RESEARCH COMPONENT

PERSPECTIVE TAKING:

EFFECTIVENESS OF TRAINING AND

IMPACT ON SENSE OF SELF

By

Sarah Berger

A thesis submitted in partial fulfillment for the degree of Clinical Psychology

Doctorate

Department of Clinical Psychology

School of Psychology

University of Birmingham

UNIVERSITYOF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

OVERVIEW

This thesis, in two volumes, is submitted as part of a three-year Doctorate of Clinical Psychology at the University of Birmingham.

Volume 1

This volume contains three papers. Paper one is a systematic literature review of research evaluating the effectiveness, teaching strategies, generalisation and maintenance of training deictic relations. Paper two is a quantitative study evaluating implicit self-referential thoughts of people with depression and their association with cognitive fusion, self-esteem and psychological distress. Paper three is an executive summary that serves as a briefing document, providing an accessible summary of the literature review and empirical paper.

Volume 2

Five Clinical Practice Reports (CPR's) are presented in this volume. Report one presents a cognitive and psychodynamic formulation of a 67-year old female, experiencing symptoms of social anxiety and panic attacks. Report two presents a service evaluation of the barriers and facilitators of a discharge process in an older adult complex care unit. Report three is a single-case experimental design evaluating the effectiveness of a cognitive behavioural assessment and intervention utilised to reduce OCD symptoms of a ten-year old female. Report four presents a

case study of a 24-year old female, identified with learning disabilities and epilepsy displaying challenging behaviour and a formulation and intervention provided from a behavioural approach. Report five is in the form of an abstract outlining a case study of a six-year old boy displaying challenging behaviour and the Dyadic Developmental Psychotherapy intervention implemented.

All names and identifying features have been changed to ensure confidentiality.

ACKNOWLEDGEMENT

I would like to pay gratitude to the following people who have provided me with their time, support and guidance.

- My research supervisor Gerry Riley, for his knowledge, patience and attention to detail. I am particularly appreciative of his ability, during our meetings together, to reinvigorate my belief that I could complete my greatest academic challenge to date.

- Coventry Mind and Kaleidoscope staff who were willing to allow me to recruit through their organisations and identified and approached participants in the first instance.

- The individuals who agreed to take part in this research who, whilst experiencing challenging times, agreed to take the time and effort to participate. Without their motivation and interest, this research would not have been possible to complete.

- Saving the most important until last, Paul Berger, my patient and supportive husband. Throughout the last three years, he has been my rock when I have doubted myself and has always been there to help me keep going. I appreciate the numerous sacrifices he has made to enable me to study and unwavering belief in me. I have no doubt that without his continuous support, the completion of this thesis would have been a far greater task. We can now finally, look forward to spending study free evenings together.

TABLE OF CONTENTS: VOLUME ONE

PAPER ONE

A SYSTEMATIC LITERATURE REVIEW OF THE EFFECTIVENESS, TEACHING STRATEGIES, GENERALISATION AND MAINTENANCE OF TRAINING DEICTIC RELATIONS

PAPER TWO

EVALUATION OF IMPLICIT SELF-REFERENTIAL THOUGHTS IN PEOPLE WITH DEPRESSION AND THEIR ASSOCIATION WITH COGNITIVE FUSION, SELF-ESTEEM AND PSYCHOLOGICAL DISTRESS

PAPER THREE

PUBLIC DISSEMINATION DOCUMENT

APPENDICES: VOLUME ONE

LIST OF TABLES: VOLUME ONE

LIST OF FIGURES: VOLUME ONE

TABLE OF CONTENTS: VOLUME TWO

CLINICAL PRACTICE REPORT 1

A PSYCHOLOGICAL MODELS CLINICAL PRACTICE REPORT

CLINICAL PRACTICE REPORT 2

A SERVICE EVALUATION OF THE BARRIERS AND FACILITATORS OF THE DISCHARGE PROCESS IN AN OLDER ADULT COMPLEX CARE UNIT

CLINICAL PRACTICE REPORT 3

A COGNITIVE BEHAVIOURAL THERAPY WITH A TEN-YEAR OLD FEMALE WITH OBSESSIVE COMPULSIVE SYMPTOMS: A SINGLE CASE EXPERIMENTAL DESIGN STUDY

CLINICAL PRACTICE REPORT 4

A CASE STUDY OF KERRY, A 24 YEAR OLD FEMALE WITH SIGNIFICANT LEARNING DISABILITIES, PRESENTING WITH CHALLENGING BEHAVIOUR: A BEHAVIOURAL APPROACH TO INTERVENTION

CLINICAL PRACTICE REPORT 5

A CASE STUDY OF KYLE, A SIX YEAR OLD BOY, SUPPORTED IN FOSTER CARE, DISPLAYING CHALLENGING BEHAVIOUR

APPENDICES: VOLUME TWO

LIST OF FIGURES: VOLUME TWO

LIST OF TABLES: VOLUME TWO

PAPER ONE

A SYSTEMATIC LITERATURE

REVIEW OF THE EFFECTIVENESS,

TEACHING STRATEGIES,

GENERALISATION AND

MAINTENANCE OF TRAINING

DEICTIC RELATIONS

ABSTRACT

Background

Deficits in perspective taking skills have been reported to have a serious impact on interpersonal communication, social cognition and result in poor functioning in social situations. Training perspective-taking skills has the potential to be an important target in clinical practice.

Aim

To evaluate the effectiveness, teaching strategies, generalisation and maintenance of training deictic relations.

Method

A systematic search was conducted for journal articles published between 2001 and 2017, to identify articles reporting on teaching deictic skills or interventions designed to encourage the use of existing deictic skills. Abstracts and titles of 988 articles were screened. Exclusion criteria were applied, resulting in 17 articles being reviewed, focusing on each study's methodology, outcome measures and their contribution to the evidence base.

2

Results

The majority of studies were of the moderate quality range. Perspective taking skills were found to meet the mastery criteria, following deictic training, but due to the poor quality of the evidence, no firm conclusion regarding the effectiveness, generalisation or maintenance of skills following training could be ascertained.

Conclusion

Methodological limitations are highlighted. The need for more high-quality research using improved teaching strategies, training conducted in naturalistic settings and follow-up data to determine long term effects is required to enable an evidence base to be further developed.

INTRODUCTION

Perspective taking has been defined as "the cognitive ability to take on another individual's viewpoint" (Baron-Cohen, Tager-Flusber & Cohen, 2000). People who have deficits in such skills, experience difficulty in trying to read the intentions of others and to comprehend the impact that their behaviour can have on those around them (Howlin, Baron-Cohen & Hadwin, 1999). The importance of perspective taking to social cognition has been recognised, with this skill enabling the ability to differentiate between mental states of the self and others, and to be able to view a situation from another's point-of-view (Hendriks et al., 2016). These skills are postulated to be precursors of more complex social behaviours, including compassion, enhanced self-awareness and feelings of empathy (Vilardaga, 2009). Therefore, deficits in perspective taking skills can have an impact on interpersonal communication, social cognition and result in poor functioning in social situations (Hendreiks et al., 2016).

Deficits in perspective taking skills have been observed in people diagnosed with schizophrenia; with these individuals scoring lower on tests such as the false-belief task (Langdon et al., 1997). Completion of this task requires the recognition of how others represent the world and how this can differ from one's own reality. Such difficulties are associated with impairments in integration and social functioning and are viewed as some of the most problematic aspects of schizophrenia (Bora, Eryavuz, Kayahan, Sungu & Veznedarolu, 2006). Furthermore, people diagnosed with autism have also been found to have deficits in perspective taking skills, in

particular, reading the emotions of others, understanding what others think of one's actions and understanding deception (Downs & Smith, 2004).

Most research on this issue has been carried out within a cognitive framework, in which the development of perspective taking is considered as a major component of acquiring 'a theory of mind' (ToM; Premack & Woodruff, 1978). Within the ToM framework, an individual is seen to progress through five stages of development. Upon the completion of the final stage, the individual is seen to have the ability to correctly predict the true or false belief of others (Baron-Cohen, 2005). An individual would be perceived to be at a less advanced stage within the ToM framework, when beliefs and motivations of others are incorrectly identified. The ToM development has been suggested to progress from simple visual perspective taking, to inferring others' beliefs and motivations and being able to predict the behaviour of others based on these states (Baron-Cohen, 2005). Developmentally, a child would typically be expected to demonstrate these skills by six years of age.

A variety of methods to assess ToM have been developed through cognitive research. However, this approach has been unsuccessful at being able to teach perspective taking and to promote the transference of these skills to other settings and tasks (Yun Chin & Bernard-Opitz, 2000). A more recent approach describing how perspective taking emerges, has been adopted by researchers within the field of behavioural analysis. This approach makes use of Relational Frame Theory, (RFT; Hayes Barnes-Holmes, & Roche, 2001; McHugh, Barnes-Holmes & Barnes-Holmes, 2004; Rehfeldt & Barnes-Holmes, 2009) which suggests that through viewing lots of examples as a child, we learn to derive relationships and eventually apply the principles independently. This theory suggests that a process is followed, where

initially a child is shown and provided with a name of an object (for example, "this is an apple"). Following this, the name (apple) is provided, and the child is able to point to the apple and then both directions of the object-name relationship are understood (object=name and name=object). Once this relationship has been taught, the child only requires the name of the object as the child knows how to determine the reverse relationship. When accomplishing this independently, the development of relational frame is seen to have occurred, with the child now being able to determine certain kinds of relationships in certain contexts, through previous training.

Nine different ways that objects and events can relate to each other have been established that have different formal attributes but share the same fundamental processes (see Table 1).

Table 1

Relationship Between Object and Event from a Relational Frame Perspective

Relationship	Example
Co-ordination	Similarity e.g. A is equal to B
Distinction	Responding to one event in relation to its differences to another event
Opposition	A is the opposite of B
Comparison	A is better/worse than B
Hierarchy	A is an attribute of B
Temporality	A happens before B
Spatiality	A is below B
Conditionality and causality	Cause-effect relationships between events "if…the…"
Perspective taking	Deictic relations

From an RFT account, deictic is a word whose reference is dependent on the context in which it is used. Perspective taking, or deictic relations, are amongst the most important part of complex cognitions, language and a sense of self, as they consist of the continual perspective from which an individual makes sense of his/her environment. Specifically, this family of relational frames include the interpersonal relations of I-YOU; the spatial relations of HERE-THERE and the temporal relations of NOW-THEN (Barnes-Holmes, Barnes-Holmes, & Cullinan, 2001). Through the analysis of people's interactions with their social environment, RFT states that taking the perspective of others can be measured (McHugh, Barnes-Holmes, Barnes-Holmes, Whelan & Stewart, 2007) and trained, through the organisation of stimuli in the social environment (Davin, Rehfeldt & Lovett, 2011). For example, perspective taking can be learnt by young children when responding to questions about both

one's own and others perspective, such as "What am *I* doing *now*?", "What did *you* notice *there*?" or "Where were *you* playing *then*?" In each case, the environment will be different when these questioned are answered, but the relational properties of I-YOU, HERE-THERE and NOW-THEN will remain constant across exemplars. Through the use of frames that define a relation in terms of perspective of the speaker, children can learn to link all sorts of different stimuli together and then generalise these to novel stimuli.

Perspective taking from an RFT context also encounters a broader meaning than in a cognitive context. This recognises the ability to distinguish perspectives of the self and a more sophisticated development of self, known as self-as-context. From this viewpoint, a continual perspective from which one experiences the internal and external environment is proposed and the ability to separate oneself from others and from one's internal and external environment is perceived to be vital to the development of the sense of self.

RFT provides a theoretical framework, suggesting that language has an impact on how we process information, which in turn influences an individual's sense of self (Foody, Barnes-Holmes, Barnes-Holmes & Luciano, 2013). An individual who attaches to the content of their thoughts, believing them to be factual, may develop a compromised sense of self, also known as 'self-as-content,' leading to a narrow range of value driven behaviour. This has been associated with a person's sense of self being located in the same relational space as the content of their thoughts (i.e. both thought and self are HERE and NOW). In contrast, an individual who is able to take an observer's perspective and view themselves as separate from the contents of their thoughts is less likely to become 'caught up' or fused with this content. This

perspective is associated with a more secure sense of self, known as 'self-as-context' (Foody et al, 2013). RFT would suggest that from this perspective, psychological content is located THERE-THEN, whilst the self is HERE-NOW, with this providing distance from thoughts and a sense of self and greater psychological flexibility.

Experimental research in this area has developed a protocol (widely known as the *Barnes-Holmes protocol*) to investigate deictic relational responses (Barnes-Holmes, 2001). This assesses deictic relations across three levels of task complexity; *simple*, *reversed* and *double reversed* (McHugh, Barnes-Holmes & Barnes-Holmes, 2004). Completing these tasks requires the participant to determine the unique relational response between stimuli in specified scenarios. For example, when responding using a *simple* relational response, no switching of perspective is required e.g. "I (experimenter) have a green pencil and you (participant) have a yellow pencil. Which pencil are you holding?" This requires an "I-YOU" relational response. When answering using a *reversed* relational response, the participant is required to answer from a changed perspective e.g. "Yesterday I was playing football, today I am swimming. If now was then and then was now, what would I be doing now?" This requires a change in perspective to answer the question from an "I-NOW" to an "I-THEN" response. At the most complex level of relational responding, known as *double reverse*, two relations are required to be reversed simultaneously, e.g. " I am sitting in a green chair and you are sitting in a red chair. If I were you and you were me and if here was there and there was here, where would I be?" This requires the participant to change perspective from an "I-HERE" to a "YOU-THERE" response. Interestingly, the correct answer for double reverse ends with "I" ending up back in the green chair. The protocol attempts to enable the participant to acquire relational

responding through progressively increasing the amount of switches of perspective taking and changing the frames that are required to be switched, using differential reinforcement and error correction over a series of trials.

Interesting patterns have been observed with deictic responding when administering the Barnes-Holmes protocol to clinical samples who have been reported to experience difficulties in perspective taking. Children diagnosed with autism spectrum disorder have been found to have lower accuracy of deictic relational tasks, in particular when responding using a reversed relational response, when compared to their typically-developing peers (Villatte et al., 2010). Additionally, individuals diagnosed with schizophrenia have been found to score lower in relational responding, when compared to a control group, with the greatest difference between groups reported during the reversal trials (Villatte et al., 2008).

A number of studies with typically-developing children have been successful in improving relational performance when using the original Barnes-Holmes protocol and feedback (McHugh 2004; Barnes-Holmes, 2002). This would suggest that deficits in deictic relational responding can be trained, and in more recent research, the same principles seem to apply within clinical populations experiencing difficulties with perspective-taking (Hendriks et al., 2016).

In summary, within both cognitive and behavioural approaches, perspective-taking is viewed as a critical basic skill that contributes to more complex behaviours and skills. These include compassion, empathy, the sense of self, emotional self-regulation, taking account of the thoughts, feelings and motivations of others in planning one's own behaviour, and taking effective part in complex social behaviours such as

humour, irony and deceit. Deficits in perspective taking skills can have a serious impact on interpersonal communication, social cognition and can result in poor functioning in social situations. They may play a role in some clinical disorders associated with social dysfunction, and, in a transdiagnostic context (such as ACT; Hayes Strosahl & Wilson, 1999), they may play a role in difficulties with emotional self-regulation. Enhancing an individual's perspective-taking skills, therefore, has the potential to be an important target in clinical practice. The cognitive approach has not yielded much in terms of how this might be done. More recent work within the context of RFT has suggested that enhancing these skills may be possible using behavioural teaching methods. The aim of this review is to critically evaluate this evidence. As part of this evaluation, the quality of the teaching methods used in these studies will also be evaluated. Behavioural teaching methods are well established and so it is possible to evaluate whether the methods used in the studies follow best practice. Furthermore, if the development of perspective-taking skills within these studies is to have practical application, it is important to evaluate the evidence about whether the skills generalise to real-life applications of perspective-taking skills and have an impact on those more complex behaviours and skills (such as emotion regulation) to which perspective-taking is suggested to contribute. Because of the limited evidence about whether the teaching of deictic skills using the Barnes-Homes protocol resulted in the transfer of these skills, the scope of the review was widened to include studies that involved prompting the use of existing deictic skills in the context of trying to alleviate some existing or experimentally-induced emotional difficulty.

11

Therefore, the current systematic literature review will aim to clarify:

1. **Is the deictic framing protocol effective in improving deictic relation performance on the task used for training?**

2. **Do the teaching strategies used in the studies follow best practice in terms of (a) improving performance on the training task, (b) improving generalisation to other settings (including everyday life) and (c) maintenance over time?**

3. **Is there evidence that teaching or prompting the use of deictic skills can generalise to real-life contexts, be maintained over time, and have an impact on more complex behaviours?**

METHOD

Search criteria were applied to the following databases: PsycINFO, MEDLINE,

EMBASE and Scopus. Publication dates were restricted to a period from 2001 to

November 2017, as journals prior to this date were not relevant due to the original

Barnes-Holmes paper that gave rise to the teaching protocol being published in 2001

(Barnes-Holmes et al., 2001). Searches were conducted using whole words, for

journals that included at least one of these key terms: *relational frame* or *deictic*

relation. The quantity of perspective taking literature necessitated the application of

limiting the search to include journals that included *relational frame* AND *perspective*

taking as this eliminated studies that included a cognitive approach to perspective

taking. Duplications were removed from the results. A four-step process was applied

to the selection process. A flow diagram is displayed in

Figure 1, outlining the selection process.
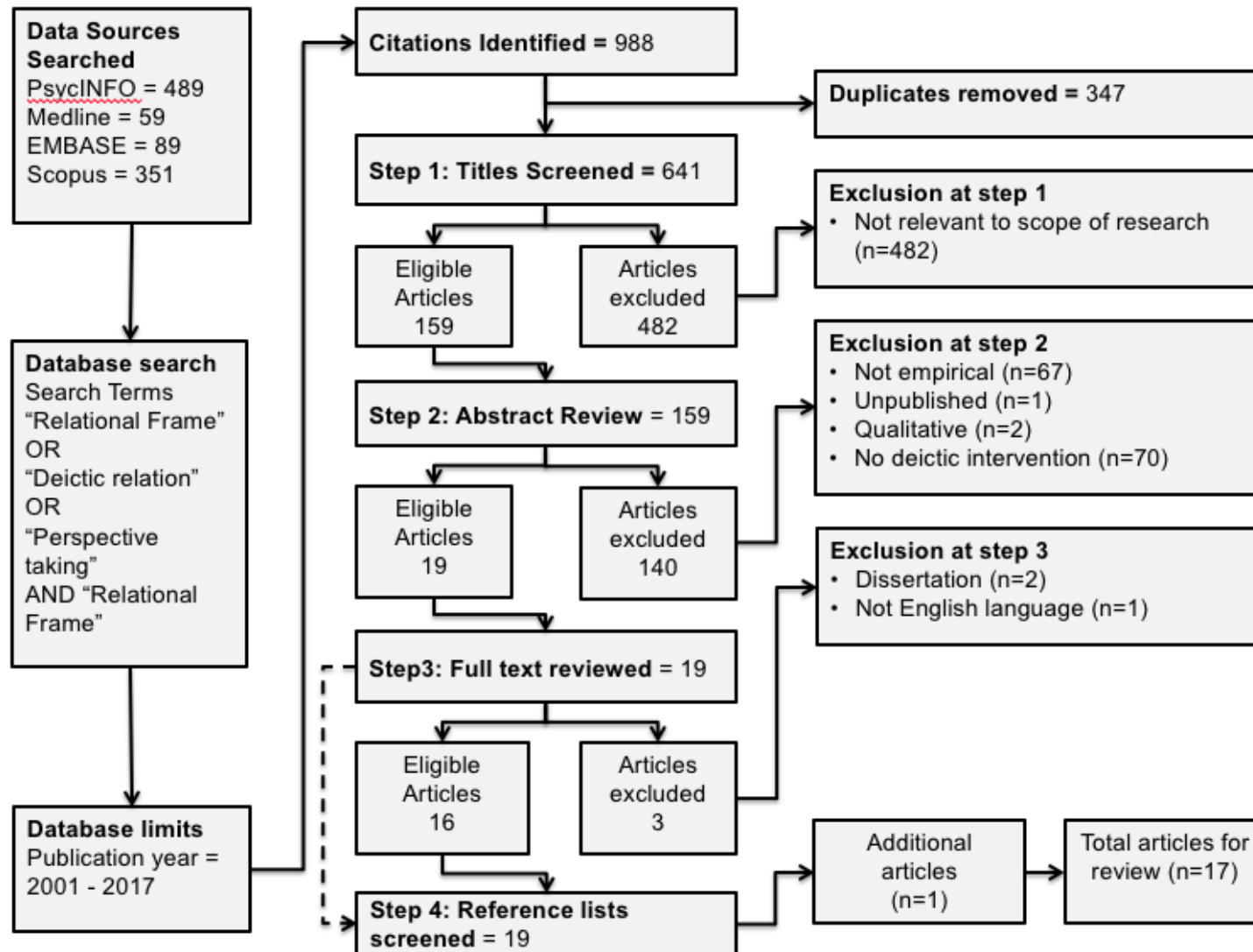
Figure 1

Flow Diagram of Article Search Strategy

Selection Process

Step 1: Results from the database search were screened to remove articles that were not related to the scope of the review. Titles of papers were initially vetted to determine whether they were relevant to the review questions.

 Step 2: Abstracts of the remaining articles were reviewed against the exclusion and inclusion criteria (see Table 2). Abstracts with limited information that did not enable the criteria to be applied were retained for further inspection.

Table 2

Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
• Involved teaching trials aimed at improving performance on tasks requiring deictic skills or an intervention designed to encourage the use of existing deictic skills	• Non-empirical papers, unpublished journals or dissertations
• Reported outcome data on whether performance improved on deictic responding and/or on some more complex skill to which deictic skills are thought to contribute	• No deictic intervention included
	• Qualitative research
• Measured a minimum of one dependent variable	• Not written in the English language

Step 3: The full text of the remaining articles were read and the inclusion and exclusion criteria were applied.

Step 4: Reference lists of the remaining eligible articles were screened for relevant studies not identified via the database search. The relevant articles located, were included in the final selection. See Table 3 and 4 for a summary of the studies included in the review.

Table 3

Summary of Studies Included in Review (Single Case Experimental Design)

	Belisle	Gilroy	Heagle, A.	Lovett	Melendez	Montoya-Rodriguez	O'Neill	Rehfeldt	Rendon	Ruiz	Weil
Year	2016	2015	2006	2014	2010	2017	2014	2007	2012	2015	2011
Location of study	USA	USA	USA	USA	Spain	USA	USA	USA	USA	USA	USA
Aim of study	Evaluated if taught perspective taking skills to adolescents with autism could transfer to novel untrained stimuli	Investigated if relational repertoires could be established using naturalistic teaching procedures in children	To improve the performance of the deictic relational protocol through training	Investigated the effect of training deictic frames to adolescents with asperser syndrome	Train perspective-taking to a three-year-old girl	Train a person with Down syndrome in deictic relational responding	Investigated if training deictic relations could facilitate theory of mind of people diagnosed with schizophrenia and mild-moderate intellectual disability	Evaluated if relational responding performance improved following reinforcement for responding relationally	Evaluated the impact of relational response reinforcement in establishing simple deictic relationships and the impact on theory of mind	To promote psychological flexibility by providing deictic and hierarchical relation training of ongoing experiences of anger	Examined the effect of operant contingencies on deictic relational responding
Variable measured	Perspective taking	Perspective taking	Theory of Mind	Perspective-taking and theory of mind	Perspective-taking	Perspective-taking	Theory of mind, level 4 and 5	Perspective-taking	Theory of mind, Perspective taking and social competence	Anger	Perspective taking and theory of mind
Participants	Children 12,14 and 18 years	Children 8,9 and 11 years	Children 6-11 years	Adults 17-18 years	Girl three years and seven months	Adult 24 years	Adults 47, 49 and 66 years	Children 9 and 10 years	Children 41-58 months	Boy five years and one month	Children 57-68 months
Sample size	3 (males diagnosed with autism)	3 (males diagnosed with autism)	3 (2 male and 1 female)	3 (2 males, 1 female) diagnosed with asperger syndrome	1 (female)	1 (male)	3 (males, diagnosed with schizophrenia and mild-moderate intellectual disability)	2 (1 male and 1 female)	4 (2 male, 2 female, raised in institutional settings)	1 (male)	3 (2 male and 1 female)

	Belisle	Gilroy	Heagle, A.	Lovett	Melendez	Montoya-Rodriguez	O'Neill	Rehfeldt	Rendon	Ruiz	Weil
Data collection method	Assess perspective taking until 100% mastery reached	Assess perspective taking until 80% mastery reached	Assess perspective taking until reached 88% simple and 90% reverse and double reverse criteria reached	Social Language Development Scenes Adolescent therapy cards Theory of mind inventory, assess perspective-taking skills until 80% mastery reached	Assess perspective taking until 100% mastery reached	Assess perspective taking until 80% mastery reached	Theory of mind level 4 and 5 testing, assess perspective taking until 80% mastery reached	Assess perspective taking until reached 90% criteria reached	Theory of mind levels 2,4 and 5, evaluation of social competence, assess perspective taking until 100% mastery reached	Functional analysis of anger and behaviour videoed and analysed	Theory of mind levels 3, 4,5. Assess perspective taking until 80% mastery reached
Design	SCED, multiple-baseline design across participants	SCED, multiple baseline across participants	SCED, multiple baseline across participants	SCED, multiple baseline across participants	SCED (A-B-A)	SCED (A-B)	SCED, multiple baseline across behaviours	SCED	SCED, multiple baseline design through participants and behaviour	SCED (A-B)	SCED, multiple baseline across participants
Multiple probe	Multiple probe (A-B-C and A-B-C-D design)	Multiple probe design across participants and relational complexity	Multiple-probe design across participants	Multiple-probe design across participant		Within-subject multiple probe design across levels of relational complexity	Within-subject multiple probe design across levels of relational complexity	Multiple-probe design across participants			Within-subject multiple probe design across levels of relational complexity
Maximum number of deictic frames to reach mastery	63	296	541	204	73	104	504	1542	24		744

	Belisle	Gilroy	Heagle, A.	Lovett	Melendez	Montoya-Rodriguez	O'Neill	Rehfeldt	Rendon	Ruiz	Weil
Deictic intervention (I-YOU, HERE-THERE, NOW-THEN)	Single -reversal deictic frame only. Picture cards presented	Five to eight sessions. Thirty-seven deictic frames (15 simple, 11 reverse and 11 double-reverse) orally presented by experimenter embedded within children's stories	Four to 8 sessions. Fifty-seven deictic frames (8 simple, 36 reverse, 13 double reverse) administered on laptop	One to three sessions per week ranging from four-six weeks. Thirty-six deictic frames trained (12 simple, 12 reverse, 12 double reverse) Pictures from therapy cards and were administered on a laptop	Eight sessions. Thirty-two deictic frames (16 simple, 16 reverse) using novel situations for each scenario Orally presented, using games and real situations	Four sessions. Reverse relation trained only Orally presented with experimenter and participant performing action	Average of 42 sessions. Fifty-six deictic frames, single, reverse and double reverse trained in sequential order. Orally presented reading from index cards	Fourteen – forty-eight sessions. Fifty-seven deictic frames, (8 simple, 36 reverse and 13 double reverse), administered on laptop	Nine sessions, six trials, simple relations only. completed blocks of four statements presented orally by experimenter	Four sessions training framing own on-going experience through deictic and hierarchical relations and then painting this on a piece of paper	15- 62 sessions. Sixty-two deictic frame scenarios simple, reverse and double-reverse relations. orally presented
Statistical analysis	None	Percentage of All Non-overlapping data	None	None	None	None	None	None	None	Tau-U-Statistics to quantify effect of intervention	None
Intervention Outcome	Bidirectional single reversal deictic framing can be improved after direct training of only single relations training. Participants demonstrated a transfer of stimulus function after both the YOU and I relations had reached mastery criteria	Participants displayed improvement of accuracy of dietic framing across all three complexity levels	Participants demonstrated the capacity to alter their perspective taking and reach criterion performance on all three relations	All three participants displayed improvement across all three levels of complexity. Varying degrees of perspective taking skills to a natural presentation to social integration based on complexity of perspective taking relation. Little change between pre and post measures of theory of mind	Perspective taking skills of simple and reverse relations showed an improvement following training using games and real situations	Perspective taking skills of reverse relations showed an improvement over trials in an adult with Downs syndrome. Improvement remained at follow-up several months later	Performance of deictic relational responding and theory of mind improved across participants	Performance of deictic responding improved following training	No clear relationship identified between improvement of simple deictic relations and scores on theory of mind and social competence	Significant reduction in duration, frequency and intensity of anger following intervention with this remaining at one-year follow-up	As improvement of the reverse and double-reverse relations was acquired, performance of theory of mind tasks measures increased
Quality rating	**Moderate**	**Moderate**	**Moderate**	**Moderate**	**Low**	**Low**	**Moderate**	**Moderate**	**Low**	**Moderate**	**Moderate**

Table 4

Summary of Studies Included in Review (Control Group Studies)

	Foody	Foody	Gil-Luciano	Jackson	Luciano	Montoya-Rodriguez
Year	2013	2015	2017	2014	2011	2016
Location of study	Ireland	Ireland	Spain	USA	Spain	Spain
Aim of study	Compared deictic vs hierarchical interventions in reducing participants' discomfort, anxiety and distress after exposure to a distress tolerance task	Analysed the effect of deictic, hierarchical, object deictic and object hierarchical interventions in promoting psychological flexibility	Analysed the effect of deictic vs hierarchical interventions in promoting psychological flexibility on a tolerance task	Evaluated the effects of increased deictic relational responding on theory of mind scores	Evaluated the impact of two defusion protocols (deictic vs deictic + hierarchical framing) with examples of promoting the function of regulating one's own behaviour	Evaluated the effect of deictic relational training on theory of mind tasks
Variable measured	Experimentally induced emotional distress (discomfort, anxiety, distress)	Distress induction task	Discomfort tolerance	Perspective taking and Theory of Mind	Problematic behaviours	Perspective taking and theory of mind
Participants	Undergraduate students age 18-21 years	Adults aged 17-41 years	Adults aged 21-46 years	Children aged 5-6 years	At risk adolescents displaying problematic behaviours age 12-15 years	Children aged 5 years
Sample size	36 (14 male, 22 female) 18 deictic training and 18 hierarchy + deictic training	48 undergraduates (22 male, 26 female), 12 deictic, 12 hierarchy, 12 object distinction, 12 object hierarchy	30 (15 males, 15 female), 10 deictic, 10 hierarchy + deictic, 10 control	5 (males) 3 experimental group, with autism, 2 control	15 (7 male, 8 female), 4 low risk-deictic, 5 low risk hierarchy + deictic, 6 high risk deictic + hierarchy	6 (males), 3 experimental, 3 control
Data collection method	Experimental Screening Questionnaire Acceptance and Action Questionnaire Visual Analogue Scale Reaction Questionnaire	Experimental Screening Questionnaire Acceptance and Action Questionnaire Visual Analogue Scale Reaction Questionnaire	Acceptance and Action Questionnaire, Cognitive Fusion Questionnaire, Depression, Anxiety and Stress Scale 21, Self-reports of pain and discomfort	Theory of mind levels 1-5, assessed perspective taking until 80% criteria reached	Behaviour Assessment Systems for Children Avoidance and Fusion, Questionnaire Accepting without Judgement Scale of the Kentucky Inventory of Mindfulness Skills, Impulsive Behaviour Inventory, Emotional Behavioural Inventory, Self-perceived utility of the protocol	Theory of mind levels 1-4, assess perspective taking until 80% mastery reached
Design	Control group study	Control group study	Control group study	Control group study	Control group study	Control group study

20

	Foody	Foody	Gil-Luciano	Jackson	Luciano	Montoya-Rodriguez
Multiple probe				Multiple probe design across relational complexity (ABACADA)		Within-subject multiple-probe design across levels of relational complexity, between-participants multiple-probe design
Maximum amount of deictic frames to reach mastery				872		636
Deictic intervention (I-YOU, HERE-THERE, NOW-THEN)	Random allocation to deictic relations (I-YOU/HERE-THERE) or hierarchical + deictic relations	Random allocation to one of four interventions; deictic relations or hierarchical + deictic relations intervention, object distinction or object hierarchy. Practice of intervention after session	Randomly assigned to control group (general interview), defusion 1 deictic relations intervention and defusion 2 deictic + hierarchical intervention + promotion of regulatory functions to that discrimination.	Nine to twenty-five sessions. Random allocation to experimental or control group. Experimental group. Sixty-two deictic frames (8 simple, 36 reverse, 18 double reverse). Orally presented by experimenter with visual prompts when required At each level of complexity completed theory of mind task was completed. Control group completed theory of mind task on four occasions across four weeks	Assigned to groups depending on low or high risk of problematic behaviours. Defusion 1 deictic relations intervention and defusion 2 deictic + hierarchical intervention function of regulating own behaviour	Four to thirteen sessions. Eighteen deictic frames reverse and double reverse with visual aids. Participants matched on deictic relation pre-test scores when split into control and experimental group. Control group completed theory of mind test on three occasions over the same time frame
Statistical analysis	Mixed ANOVA utility of interventions of anxiety, discomfort and stress. Dependent t-test mean scores of DV's	Mixed ANOVA - effects of intervention on anxiety, discomfort and distress, dependent t-test mean scores of DV's	Mixed ANOVA - explored equivalence of experimental conditions	None	Wilcoxon Z and Mann-Whitney U to measure pre-post protocol	None

	Foody	Foody	Gil-Luciano	Jackson	Luciano	Montoya-Rodriguez
Intervention Outcome	Following deictic intervention, significant reduction in believability and less distraction. Hierarchical deictic intervention superior to deictic alone relation in reducing distress. Little difference interventions between discomfort, anxiety	Following deictic intervention significant reduction in stress. Hierarchy was superior to reduce stress compared to deictic but two groups not significantly different. Little difference between deictic and hierarchy interventions on reducing discomfort and anxiety	Participants who received defusion 1 protocol of deictic framing performed significantly better in pain tolerance during the cold compressor task post-test, compared to control. Defusion 2 protocol of deictic framing + hierarchical framing + promotion of regulatory functions to that discrimination displayed significantly greater tolerance to pain compared to deictic alone following intervention	Deictic relational responding for the experimental group was found to improve with training but this was not sufficient to change more generalised perspective taking skills as measured by theory of mind. Control group who repeated theory of mind task only, did not display a consistent increase in scores	Deictic framing intervention- no significant reduction problematic behaviours or psychological flexibility. Deictic + hierarchical framing + promoting regulation intervention displayed significant reduction of problematic behaviours with this remaining at 4 months follow-up.	Deictic relational responding improved in the experimental group. No difference in theory of mind scores between experimental and control group
Quality rating	**High**	**Moderate**	**Moderate**	**Moderate**	**High**	**Low**

Quality evaluation

The quality of studies under review was assessed to determine potential methodological bias. This was important as it enabled judgements to be made about the confidence one could place on each study's findings. This process was found to be useful as there were conflicting results across studies. However, it was acknowledged that there was limited empirical evidence supporting the reliability and validity of applying quality criteria in practice and that, due to this, caution regarding the judgments made should be employed (Higgins & Green, 2006).

The Risk of Bias in N-of-1 Trials (RoBint; Tate et al., 2015) was used, as this was a tool designed to facilitate the critical appraisal of single-case intervention studies. It was deemed to be a good instrument to use due to it providing a "practical, valid, sensitive and reliable scale" (Tate et al., 2015). This 15-item rating scale enabled the assessment of both internal and external validity. Each of the 15 items was awarded a maximum of two points. These were combined to provide an overall score of methodological quality out of a maximum of 30 points. This was further broken down into internal validity (maximum of 14 points) and external validity (maximum of 16 points). Descriptive categories were applied to studies by their overall score quality rating. Scores ranging from 0-10 were rated as low, 11-20 were rated medium and 21-30 were rated as high. For full details of the quality appraisal checklist see Appendix 1.

A "quality appraisal checklist for intervention studies" provided by that National Institute of Clinical Excellence was used for the quantitative studies that were not single-case interventions (NICE, 2012). This measure included five sections; Section

1 assessed the key population criteria for determining external validity, sections 2-4

assessed whether the outcomes reported were attributable to the intervention rather

than any confounding factors and section 5 assessed bias and whether findings

reported were generalisable. For full details of the quality appraisal checklist see

Appendix 2. These five sections were scored from a list of five responses (see Table

5).

Table 5

Appraisal Checklist Scoring Responses

Scoring Symbol	Full description
++	Indicated that the specific aspect of study design being assessed, had been designed or conducted to minimise the risk of bias
+	Indicated that either the answer to the checklist question was not clear from the way the study was reported, or that the study may not have addressed all potential sources of bias for that particular aspect of study design
-	Reserved for aspects of the study design in which significant sources of bias could persist
Not reported (NR)	Reserved for those aspects in which the study under review failed to report how they had (or might have) been considered
Not applicable (NA)	Reserved for study design aspects that were not applicable given the study design under review

An overall quality grading was awarded to each study for internal validity and external

validity. This was determined by assessing each study against a quality grading

criteria (see Table 6).

Table 6

Overall Quality Grading

Quality Grading	Internal/External Validity Description
++	All or the majority of the checklist criteria had been met, where they had not been met the conclusions were very unlikely to alter
+	Some of the checklist criteria had been met, where they had not been met, or not sufficiently described, the conclusions are unlikely to alter
-	Few or no checklist criteria had been met and the conclusions were likely or very likely to alter

Key information was extracted for each of the studies against one of the quality criteria. Table 7 and Table 8 display the quality scoring for each study scored.

Table 7

Quality of the Single-Case Report Studies Included in the Review Using the RoBINT Scale (Tate et al., 2015)

	Belisle, J.	Gilroy, S.P.	Heagle, A.I.	Lovett, S.	Melendez, G.	Montoya-Rodriguez	O'Neill, J.	Rehfeldt, R.A.	Rendon, M.I.	Ruiz, F.	Weil, T.M.
Publication date	2016	2015	2006	2014	2010	2017	2014	2007	2012	2014	2011
Design with control	2 (multiple probe x3)	2 (multiple probe x3)	2 (multiple probe x3)	2 (multiple probe x3)	1 (A-B-A)	0 (A-B)	2 (multiple probe x3)	1 (2 A_B studies)	2 (multiple probe x4)	0 (A_B design)	2 (multiple probe x3)
Randomis-ation	0 (not mentioned)	0 (based on performance)	0 (not mentioned)	0 (not mentioned)	0 (one participant)	0 (one participant)	0 (not mentioned)	0 (not used)	0 (no randomisation other than baseline measure)	0	0 (no randomisation mentioned)
Sampling of behaviour	0 (less than three points at baseline)	1 (4 data points at baseline)	0 (less than 3 data points at baseline)	0 (less than three data points at baseline)	0 (less than 3 data points at baseline)	0 (less that 3 data points at baseline)	0 (less than 3 data points at baseline for simple relation)	0 (less than three data points at baseline)	0 (less than three points at baseline)	2 (5+ baseline and intervention)	0 (less than 3 data points at baseline)
Blinding involved in people in the intervention	0 (no blinding mentioned)	0 (no blinding mentioned)	1 (computer generated in random order, but participant knows which level)	1 (computer generated in random order, but participant knows which level)	0 (blinding not mentioned	0 (blinding not mentioned	0 (no blinding mentioned)	1 (computer generated in random order, but participant knows which level)	0 (blinding not mentioned	0 (blinding not mentioned	0 (blinding not mentioned
Blinding of assessor	0 (no mention of blinding)	0 (not blinded)	2 (assessed using a computer)	2 (assessed using a computer)	0 (read out instructions)	0(read out instructions)	0 (experimenter reads out questions, no info on who scored data)	2 (computer generated questions)	0 (experimenter reads out questions, no info on who scored data)	0 (no info on who scored data)	1 (experimenter reads out questions, but observer scored data)

	Belisle, J.	Gilroy, S.P.	Heagle, A.I.	Lovett, S.	Melendez, G.	Montoya-Rodriguez	O'Neill, J.	Rehfeldt, R.A.	Rendon, M.I.	Ruiz, F.	Weil, T.M.
Interater-agreement	2 (32% of trials, agreement 100%)	2 (34%, all sessions, 98.9%)	2 (computer program and generalisation assessed, 100%)	2 (computer program and generalisation assessed, 100%)	0 (no inter-rater agreement mentioned)	2(100% of session observed and 95%)	2 (33% of all sessions, 95.8% agreement and each condition included)	2 (computer program generated)	0 (no inter-rater agreement mentioned)	0 (no inter-rater agreement mentioned)	2 (33% of sessions, observer across all sessions rated and 90%)
Treatment adherence	0 (nothing about who assessed the data to determine if independent assessor)	2 (checklist used, every session, 100%)	2 (use of computer to implement)	2 (use of computer to implement)	0 (no evaluation of treatment adherence)	0 (no evaluation of treatment adherence)	0 (nothing about who assessed the data to determine if independent assessor)	2 (use of computer to implement)	0 (nothing about who assessed the data to determine if independent assessor)	0 (no evaluation of treatment adherence)	2 (observer rated, clear criteria, 33% of data rated, 93% integrity)
Baseline characteristics	2 (age, sex, IQ scores, PEAK-D scores for relational ability)	1 (low observed accuracy of deictic - not specifically measured)	1 (reading but no specific info measured)	0 (no details for each individual participant and no measure measured - only used school records)	0 (no indication of the level of the behaviour severity at baseline - just under 100%)	0 (no aetiology and no level of baseline behaviour)	1 (info on baseline condition re schizophrenia deictic, anhedonia, 80% or less deictic + ages, sex, aetiology but no info of LD)	0 (no info on condition, factors maintain behaviour being measured, baseline behaviour not measured/ reported)	0 (no severity of condition of baseline data for participants provided)	1 (vague description of relational responding pre-intervention)	1 (no info on educational background)
Setting	2 (room and description of setting provided)	2 (room of participant and description of setting)	1 (limited info regarding environment of room)	2 (room and description of setting provided)	1 (conducted in a playroom but no description of environment)	2 (setting and environment described)	1 (vague info of setting and environment)	1 (no info about the room)	1 ("space in the foundation" but no description of environment)	1 (info re setting at participants home but no description)	2 (setting and environment described)

	Belisle, J.	Gilroy, S.P.	Heagle, A.I.	Lovett, S.	Melendez, G.	Montoya-Rodriguez	O'Neill, J.	Rehfeldt, R.A.	Rendon, M.I.	Ruiz, F.	Weil, T.M.
Dependent variables	1 (behaviour defined but vague)	2 (accuracy of relational responding and measured dividing no of correct responses over incorrect)	2 (relational frames described and how measured specified - 88%)	2 (relational frame described and how measured specified)	1 (process described but behaviour not described well)	1 (process described but behaviour not described well)	2 (relational frames described and how measured specified - 80%)	1 (method described but target behaviour unclear)	1 (process described but target behaviour not described well)	2 (anger described, and measures reported)	2 (relational frames described and how measured specified - 80%)
Independent variable	0 (no frequency, duration, provided)	2 (amount of sessions, duration and frequency provided, described)	1 (sessions not specific described - "at least three sessions administered three times a week"	0 (vague description, 30-45 mins, one -three times a week, ranged from four to six weeks and no of sessions not provided)	2 (description of intervention and procedure of delivery both included)	1 (description provided but no frequency of sessions)	2 (sessions on graph, duration and frequency provided)	0 (no frequency, duration, provided)	1 (no frequency provided)	2 (description of intervention and procedure of delivery both included)	0 (no frequency for p3 and no duration provided)
Raw data record	2 (each data point on graph provided)	2 (each data point provided)	2 (all data points provided visually)	2 (each data point provided)	2 (all data points provided visually)	2 (all data points provided visually)	2 (all data points provided visually)	0 (no baseline data provided)	2 (all data points provided visually)	2 (all data points provided visually)	1 (aggregated data provided)
Data analysis	0 (no data analysis)	1 (stats used but no justification)	0 (visual inspection of trend)	0 (visual inspection of trend)	0 (no mention of visual inspection or analysis of data)	0 (no mention of visual inspection or analysis of data)	0 (no mention of visual inspection or analysis of data)	0 (no data analysis)	0 (only visual inspection)	2 (Tau-U-statistic used, and justification provided)	0 (no mention of visual analysis)
Replication	1 (only 3 cases)	1 (only 3 cases)	1 (only 3 cases)	1 (only 3 cases)	0 (only 1 case)	0 (only 1 case)	1 (only 3 cases)	1 (2 cases)	1 (4 cases)	0 (only 1 case)	1 (3 cases)

	Belisle, J.	Gilroy, S.P.	Heagle, A.I.	Lovett, S.	Melendez, G.	Montoya-Rodriguez	O'Neill, J.	Rehfeldt, R.A.	Rendon, M.I.	Ruiz, F.	Weil, T.M.
General-isation	1 (transfer between phases but vague)	0 (only measured at the end of the intervention phase)	2 (after each phase and scripted)	2 (after each phase video recording and written responses)	0 (not mentioned)	0 (not mentioned)	1 (pre and post probes and generalisation mentioned in method and discussion)	0 (not mentioned)	1 (pre and post probes and generalisation mentioned in method and discussion)	0 (not mentioned)	1 (pre and post probes and generalisation mentioned in method and discussion)
TOTAL SCORE	13	18	18	18	7	8	14	12	9	12	15

Key:

■ = studies of medium quality

■ = studies of low quality

Table 8

Quality of the Experimental Design Studies Included in the Review Using the Quality Appraisal Checklist (NICE, 2012)

Quality Criteria assessed	Foody (2013)	Gil-Luciano (2017)	Jackson (2014)	Foody (2015)	Luciano (2011)	Montoya-Rodriguez (2016)
Well described population	+	+	+	+	+	+
Eligible population	N/R	N/R	N/R	NR	N/R	N/R
Representative population	N/R	N/R	N/R	NR	N/R	N/R
Overall external validity	N/R	N/R	N/R	NR	N/R	N/R
Allocation to intervention	+	+	+	+	-	+
Description of intervention	++	+	+	++	+	+
Allocation concealed	N/R	NR	N/R	NR	-	-
Blinding or participants and experimenter	+	-	-	-	-	-
Adequate exposure to intervention	+	++	+	+	+	++
Contamination level low	+	+	++	+	+	++
Similarity of interventions	++	++	+	++	++	+
Dropout rates/follow-up	++	++	++	++	++	++
Setting reflect usual UK practice	+	+	+	+	N/A	N/A
Intervention reflect usual UK practice	+	+	N/R	+	+	N/R
Reliable outcome measures	+	-	+	+	++	+
Complete outcome measures	++	++	++	++	++	++
Important outcomes assessed	N/A	N/A	N/A	NA	N/A	N/A
Relevant outcomes included	N/A	N/A	N/A	NA	N/A	N/A
Follow-up times in exposure and control group	-	-	-	+	++	-
Follow-up time	N/A	N/A	N/A	NA	+	N/A
Baseline	++	+	+	+	++	+
Intention to treat	N/A	N/A	N/A	NA	N/A	N/A
Sufficient power	-	N/R	N/R	NR	-	-
Estimates of effect size	++	++	N/R	++	++	N/R
Analytical methods	++	++	+	++	++	-
Confidence intervals reported	++	++	N/R	++	++	-
Internal validity	+	+	+	+	+	+
Generalisable	-	-	-	-	-	+
Overall rating	++	+	+	+	++	-

Key: N/R = not recorded; N/A - Not applicable; ++ = Design and study minimise the risk of bias; + = Study or design may not have addressed all potential sources of bias; - = Significant source of bias may persist

As noted earlier, another aim of the review was to evaluate whether the teaching strategies used in the studies followed best practice recommendations in terms of; (a) improving performance on the training task and (b) improving generalisation to other settings (including everyday life) and transfer to other tasks? To address this aim, another evaluative framework was developed on the basis of literature about effective behavioural teaching strategies (Martin & Pear, 1992; Miltenberger, 2000; Morse & Schuster, 2004; Stokes & Baer, 1977). The framework evaluated the effectiveness of the teaching strategy in terms of how well it engaged the participants, the use of reinforcement, the use of prompts and feedback and the promotion of the generalisation and maintenance of the taught responses. Table 9 provides the ratings on individual items for the reviewed papers. No overall score was given to individual papers. Instead, the evaluative tool was used to identify areas that were not particularly well addressed by the studies in general.

Table 9

Quality of Teaching Across Studies

	Belisle, J.	Gilroy, S.P.	Heagle, A.	Jackson, M.L.	Lovett, S.	Melendez, G.	Montoya-Rodriguez, M.M.	Montoya-Rodriguez, M.M.	O'Neill, J.	Rehfeldt,R. A.	Rendon, M.I.	Weil, T.M.
Year published	2016	2015	2006	2014	2014	2010	2016	2017	2014	2007	2012	2011
Assessment of task engagement	no, low interest	no, but did determine "program end point to reduce frustration of two sessions without improvement	no, low level of interest	no, low level of interest	no, low level of interest	no, low level interest	no , low level of interest	no, low level of interest	no , low level of interest	no, low level of interest	no, but did alternate between work and play	no, low level of interest
Assessment of teaching materials of interest to participants	no, used 6 picture cards	no, but materials likely to be of interest	no, low level of interest as repetitive	no, low level of interest as repetitive	no, but used stimuli that was specific for adolescents so could be greater engagement	no, but materials likely to be of interest	no, but used new scenarios that were different to increase level of interest	no, but materials likely to be of interest	no, low level of interest as repetitive	no, low level of interest as repetitive	no, low level of interest as repetitive	no, low level of interest as repetitive
Measures taken to enhance task engagement and avoid boredom?	no	yes, popular Disney story books used. Scheduled breaks provided if required	no, but brief breaks were provided when required where played games with experimenter	no, but breaks every 20 mins Pleasurable activity of child's choice after each session	no, but short breaks offered every 15 mins	yes, used toys of the chilld	no, but pleasurable activity of child's choice after each session	yes, used coloured pencil, paper, scissors	no, sessions were kept to 10 mins, so boredom may have been reduced	no, but brief breaks were provided when required where played games with experimenter	no, no mention of breaks	no
Correct response intrinsically reinforcing?	no	no	no	no	no	no	no	no	no	no	no	no
Assessment that consequences were reinforcing to participant?	no	yes , preference assessment with carer and participant	no	yes, preference assessment conducted and allowed brief access	no	no	no	no	yes, self-report measures of preferred items	no	no	yes, preference assessment with carer and participant

32

	Belisle, J.	Gilroy, S.P.	Heagle, A.	Jackson, M.L.	Lovett, S.	Melendez, G.	Montoya-Rodriguez, M.M.	Montoya-Rodriguez, M.M.	O'Neill, J.	Rehfeldt,R.A.	Rendon, M.I.	Weil, T.M.
Reinforcement delivered immediately after correct response?	yes, answer correctly within 3 seconds = verbal praise	? say reinforced but not stated how or when this was. Provided at the end of the trial stated	yes, 3 second animation clip	? after two questions answered correctly and doesn't state what this looked like	yes, "correct" appeared on screen	no	yes "great job"	yes	yes "praise" offered after 2 questions correctly answered	yes, 3 second animation clip	yes, verbally reinforced	yes, after second question answered correctly,
Verbal praise provided	yes, but doesn't state what "verbal praise" was	yes, but doesn't state what "verbal praise" was	no	? Feedback provided but not specified what was said	no "correct" appeared on screen	no	yes "great job" for correct answer	yes, verbal praise	yes "yes that's correct"	no, animation appeared on screen	yes- "very good" after correct response	yes, but doesn't state what "verbal praise" was
Feedback given about whether response was correct	yes, answer correctly within 3 seconds = verbal praise	? states provided feedback after correcting an error but nothing about correct answers generally	? 3 second animation clip provided but doesn't state what this looked like	? states "correct responses were reinforced" but doesn't say what this looked like	yes "correct" appeared on screen	no	yes "No, I'm sorry, I would be in correct answer"	yes, verbally provided	yes, verbally provided for correct and incorrect responses	yes, 3 second animation clip	yes, verbally provided for incorrect and cyrrect responses	yes, after second question correctly answered gain social praise
Clear procedure for eliciting correct response in the event of error	yes "try again" and then provided the answer	yes "no, say….." if correct, praise provided.	? "try again" and trial repeated. Not specific feedback	? states "incorrect responses resulted in corrective feedback " but doesn't say what this looked like	? "try again" on screen and trial represented until answered correctly	yes, repeat contextual key and physically pointing it out add repeating key	yes "No, I'm sorry, I would be (correct answer) and then trial repeated	yes " no I'm sorry, I would cut the paper person"	yes "no I would have the red shirt"	? "try again" and trial repeated. Not specific	yes, had 3 levels of progression of training if 100% not reached on each level	yes "no, I'm sorry, I would be in the blue chair"
Prompts used to increase correct response	yes, picture cards	no	no	yes, visual prompts. Participant and researcher would hold picture cards	no	no	yes, researcher and participant placed themselves in the others place (reversal)	yes, paper with names was provided to prompt correct questions	no	no	yes, cards with drawings on provided and actions and spaces provided additionally	no

33

	Belisle, J.	Gilroy, S.P.	Heagle, A.	Jackson, M.L.	Lovett, S.	Melendez, G.	Montoya-Rodriguez, M.M.	Montoya-Rodriguez, M.M.	O'Neill, J.	Rehfeldt, R.A.	Rendon, M.I.	Weil, T.M.
Clear strategy for withdrawal of prompts	no	no	no	yes, once participant reached 80% criteria , prompts removed	no	no	yes, gradually reduced the physical movement into the position of the other person until relation was presented only verbally	no	no	no	no	no
Teaching take place in participants' everyday environment when opportunities naturally arose that required deictic skills	yes, in class-room	yes, home environment	yes, university setting	yes, room in their home or research room	no experiment room	yes, in a playroom where the stimuli would usually be placed but material not naturally occurring	yes, children's school and completed training in the context of activities that were carried out	no, day-care centre	? -home & adult day services but not naturally occurring	no - rehab centre	no, not specified	yes, room at home, university, school environment
Assessment to see if acquired skills-maintained overtime?	no	No	no	no	no	no	no	yes, at follow-up several months later	no	no	no	no
If follow-up measure taken, was this after two months?	no	No	no	no	no	no	no	yes - "several months"	no	no	no	no
Assessment of generalisation of the skills to a different context?	yes, stimuli changed, different question	yes, different experimenter	yes, different stimuli and conversation rather than laptop format	no	yes, videoed scenarios post phase	no	yes, stimuli changed, different questions	no	no	no	yes, stimuli changed, different questions	yes, stimuli changed, different questions
Assessment of whether the skills generalised to everyday life?	no	no	no	no	yes, scenarios of situations that could occur in student's daily life	no	no	no	no	no	no	no
Assessment of whether skills generalised to another task?	no	no	no	yes, tested TOM but no change found	yes, TOM, social language skills and perspective taking	no	yes, TOM	no	yes, TOM	no	yes, TOM and social competence	yes, TOM

34

	Belisle, J.	Gilroy, S.P.	Heagle, A.	Jackson, M.L.	Lovett, S.	Melendez, G.	Montoya-Rodriguez, M.M.	Montoya-Rodriguez, M.M.	O'Neill, J.	Rehfeldt,R. A.	Rendon, M.I.	Weil, T.M.
Multiple exemplars used to teach the skill in order to enhance generalisation	No	no, only for post phase	no, only for post phase	no	yes, variation of task material was used for generalisation with video and answered on paper	no	no	no, only for post phase	no, only for post phase	no	no, only for post phase	no, only for post phase

Key: ? = findings not clear

RESULTS

Summary of Evaluations

Quality

The eleven single case intervention studies included in the review were found to be predominantly of moderate quality range when assessed using The Risk of Bias in N-of-1 Trials Scale (RoBint; Tate et al., 2015) (see Table 7). Three studies scored within the low range, eight in the moderate range and no studies were found to be within the higher quality range. Limitations of the studies included nine of the eleven studies not reporting any data analysis, making it difficult to determine the significance of the outcomes reported. Only one study, (Ruiz & Perete, 2015), included five data points at baseline, allowing for the stability of the baseline to be assessed adequately and six of the studies did not report evidence for treatment adherence. However, strengths of the studies included seven of the studies using a multiple-baseline design and eight studies also providing inter-rater agreement to ensure consistency of ratings of the target behaviour being assessed.

Six experimental design studies included in the review were all found to have low levels of external validity. None of these studies reported whether the sample was presentative of the population or representative of the source (see Table 8). Two studies scored within the higher range when assessing internal validity, suggesting that the outcome of these measures were valid. Three studies fell into the moderate range, suggesting that whilst the outcome of these studies was unlikely to alter, not

all sources of bias had been addressed and one fell into the low range suggesting conclusions were likely or very likely to alter.

Location

The studies included were all conducted in developed westernised countries; United States of America (*n*=11), Spain (*n*= 4) and Ireland (*n* =2).

Age

The age of participants ranged from the youngest participant being 3 years and 7 months (Melendez, 2010) to 66 years of age (O'Neill & Weil, 2014).

Settings

The settings where the interventions were carried out were predominantly within the home (*n*= 6) (Gilroy, Lorah, Dodge & Fiorello, 2015; Jackson, Mendoza & Adams, 2014; Melendez, 2010; O'Neill & Weil; 2014; Ruiz & Perete, 2015; Weil, Hayes & Capurro, 2011) and educational settings (*n* = 5) (Belisle, Dixon, Stanley, Munoz & Daar, 2016; Foody et al., 2013; Foody, Barnes-Holmes, Barnes-Holmes, Rai & Luciano, 2015; Montoya-Rodriguez, McHugh & Cobos, 2017; Rehfeldt, Dillen, Ziomek & Kowalchuk, 2007). The remainder took place in either an "experimental room" (*n*= 3) (Gil-Luciano et al., 2017; Lovett & Rehfeldt, 2014; Luciano et al., 2011), day-care centre (*n* = 1), (Montoya-Rodriguez et al., 2017) rehabilitation institute (*n*-= 1), (Heagle & Rehfeldt, 2006) or the setting was not specified (*n*= 1) (Rendon, Soler & Cortez, 2012).

Characteristics of included studies

The sample sizes ranged from ($n = 1$) (Melendez, 2010; Montoya-Rodriguez & Cobus, 2016; Ruiz & Perete, 2015) to ($n = 48$) (Foody et al., 2015) with a mean number of participants across all studies of 10. The design of the studies included eleven single-case experimental designs (Belisle et al., 2016; Gilroy et al., 2015; Heagle & Rehfeldt, 2006; Lovett & Rehfeldt, 2014; Melendez, 2010; Montoya-Rodriguez et al., 2017; O'Neill & Weil, 2014; Rehfeldt et al., 2007; Rendon et al., 2012; Ruiz & Perete, 2015; Weil, Hayes & Capurro, 2011) and six control group studies (Foody et al., 2013; Foody et al., 2015; Gil-Luciano et al., 2017; Jackson et al., 2014; Luciano et al., 2011; Montoya-Rodriguez & Cobos, 2016). The participants in the majority of the studies were typically developing children/adults ($n = 10$) (Foody et al., 2013; Foody et al., 2015; Gil-Luciano et al.,2017; Heagle & Rehfeldt, 2006; Melendez, 2010; Jackson et al., 2014; Montoya-Rodriguez & Cobos, 2016; Rehfeldt et al., 2007; Rendon et al., 2012; Weil,et al., 2011). Four of the studies recruited individuals with developmental disorders, which included participants with a diagnosis of autism/aspergers (n=3) (Belisle et al., 2016; Gilroy et al., 2015; Lovett & Rehfeldt, 2014) and Downs Syndrome ($n = 1$) (Montoya-Rodriguez et al., 2017). Three further studies utilised a clinical sample, that included schizophrenia and mild to moderate intellectual disability ($n = 1$) (O'Neill & Weil, 2014) and anger and problematic behaviour ($n = 2$) (Ruiz & Perete, 2015; Luciano,et al., 2011).

1. Is 'Barnes-Holmes' protocol effective in improving deictic relation performance on the task used for training?

Twelve papers, two controlled group studies and ten SCED's, assessed whether perspective taking skills could be enhanced through deictic relational training, using I-YOU, HERE-THERE, NOW-THEN (4 low, 8 moderate quality).

All 12 papers reported that perspective training skills improved to meet mastery criteria specified after training, with seven studies setting 80%, two 90% and three studies 100% mastery criteria. Seven studies trained all three levels of complexity (Gilroy et al., 2015; Heagle & Rehfeldt, 2006; Jackson et al., 2014; Lovett & Rehfeldt, 2014; O'Neill & Weil, 2014; Rehfeldt et al., 2007; Weil et al., 2011), two trained two relations (Melendez, 2010; Montoya-Rodriguez & Cobos, 2016) and three, only one relational frame (Belisle et al., 2016; Rendon et al., 2012; Montoya-Rodriguez et al., 2017).

All studies used the Barnes-Holmes protocol, but the administration of presenting the training varied. Methods included the use of a laptop (Heagle & Rehfeldt, 2006; Lovett & Rehfeldt, 2014; Rehfeldt et al., 2007), verbal instructions provided orally (Jackson et al., 2014; O'Neill & Weil, 2014; Rendon et al., 2012; Weil et al., 2011), picture cards (Belisle et al., 2016), activity format (Montoya-Rodriguez & Cobos, 2016; Montoya-Rodriguez et al., 2017), framing through children's stories (Gilroy et al., 2015) and including a game element (Melendez, 2010). The results would suggest that the administration of the protocol can be applied in a variety of ways and still remain effective.

The number of deictic frames required to reach mastery criteria of the protocol across the twelve studies ranged from 24 simple frame relations only (Rendon et al., 2012) to 1542 relations, covering three levels of complexity (Rehfeldt et al., 2007). These results would suggest that even if training deictic relations is effective, a large amount of time and effort may be required to reach mastery criteria.

Mastery of the protocol was achieved across a variety of populations. One low quality study reported improvement with an adult with Down syndrome (Montoya-Rodriguez et al., 2017), four moderate quality studies found improvement in people diagnosed with autism (Belisle et al., 2016; Gilroy et al., 2015; Jackson et al., 2014; Lovett & Rehfeldt, 2014) and one moderate quality study reported an increase in scores of perspective taking with adults diagnosed with schizophrenia and mild to moderate learning disability (O'Neill & Weil, 2014). Additionally, improvements in typically developing participants were reported in three low quality studies (Melendez, 2010; Montoya-Rodriguez & Cobos, 2016; Rendon et al., 2012) and three moderate quality studies (Heagle & Rehfeldt, 2006; Rehfeldt et al., 2007; Weil et al., 2011).

Limitations of the studies that incorporated the protocol included only one moderate quality study, using statistical analysis to assess improvements following intervention (Gilroy at al. 2015). The remaining studies were reliant on visual inspection of the data alone, increasing the possibility of making a type I error, and inferring that a significant change had occurred following the intervention, when this may not be an accurate result (Crosbie, 1993). Only two controlled group studies compared the effectiveness of improving deictic relations compared to a control group (Jackson et al., 2014; Montoya-Rodriguez & Cobos, 2016), with both of these studies not using statistical analysis to determine if change was significant between groups. All other

studies were SCED's which utilised a small sample size, therefore not allowing for the generalisation of findings to the population. The design of all studies utilising the protocol had limited data points at baseline phase. Therefore, the variability and baseline trend were unable to be established, reducing the confidence that the effect observed could be linked to the deictic training.

Although mastery criteria were reached in all of the studies, the quality of the evidence was poor and one cannot conclude with any certainty whether the Barnes-Holmes protocol is effective in improving deictic skills.

2. Do the teaching strategies used in the studies follow best practice in terms of (a) improving performance on the training task (b) improving generalisation to other settings (including everyday life) and (c) maintenance over time?

The first issue evaluated using the best practice framework (Appendix 3), concerned how well the studies assessed and promoted the engagement of the participants in learning the task. The protocol required a considerable number of trials and sessions across many of the studies with, for example, Belisle and colleagues (2016) reporting 63 trial blocks to complete the task. This would be hypothesised to increase levels of frustration and boredom and reduce the level of interest for some participants, with two studies needing to adapt their training to incorporate scheduled breaks and a token economy to improve the engagement (Gilroy et al., 2015; Weil et al., 2011). None of the studies that used the Barnes-Holmes protocol conducted an assessment of task engagement or an assessment of whether the teaching material was of interest. The failure to deal effectively with engagement and reinforcement issues

may have resulted in the considerable amount of teaching required in some of the

studies to achieve the mastery criterion. However, three of the studies (2 low, 1

moderate quality) did take some measures to try and enhance the engagement of

participants, by presenting the deictic frames through the use of toys, Disney stories

and drawing (Gilroy et al., 2015; Melendez, 2010; Montoya-Rodriguez et al., 2017).

However, as the level of interest was not measured, the impact of this is unknown.

Nevertheless, tailoring the intervention to the audience may well have improved the

effectiveness of the teaching strategy when training deictic relations, with research

proposing implementing this to have a positive impact on intervention outcomes and

engagement (Yardley et al., 2016).

Remaining engaged on the task was found challenging for some young children

diagnosed with autism and typically developing young children. Duration of sessions

ranged from 10 to 60 minutes, with five studies providing breaks throughout sessions

(Gilroy et al., 2015; Heagle & Rehfeldt, 2006; Jackson et al., 2014; Lovett & Rehfeldt,

2014; Rehfeldt et al., 2007). Overall, there are serious questions about how well the

training in these studies engaged the participants.

Providing a correct response whilst completing the protocol was not intrinsically

reinforcing in any of the studies, with the correct perspective taking response not

leading to reinforcing consequences, if used in the participant's everyday natural

environment. Four studies did try to ascertain whether artificial reinforcement would

provide positive reinforcement of correct responses by including a preference

assessment (Gilroy et al., 2015; Jackson et al., 2014; O'Neill & Weil, 2014; Weil et

al., 2011). In other cases, however, it was not clear how reinforcing task completion

was for the participants. Four studies provided pleasurable activities at the end of

each session in an attempt to keep participants on task (Heagle & Rehfeldt, 2006; Jackson et al., 2014; Montoya-Rodriguez & Cobos, 2016; Rehfeldt et al., 2007). Eight studies provided immediate reinforcement after a correct response in the form of a three second video clip (Heagle & Rehfeldt; 2006; Rehfeldt et al., 2007), verbal praise, (Belisle et al., 2016; Gilroy et al., 2015; Montoya-Rodrigues et al., 2016; Montoya-Rodrigues et al., 2017; Rendon et al., 2012;) or "correct" appearing on the screen (Lovett & Rehfeldt, 2014). Three studies provided positive reinforcement after two consecutive responses (Jackson et al., 2014; O'Neill & Weil, 2014; Weil et al., 2011).

Another aspect of assessing whether the studies followed best practice in terms of improving performance on the training task concerned their use of prompts. A clear procedure for eliciting the correct response in the event of an error was provided in eight of the studies, and five studies included prompts to increase the probability of responding correctly using visual prompts or physically completing the action (Belisle et al., 2016; Jackson et al., 2014; Montoya-Rodriguez & Cobos, 2016; Montoya-Rodriguez et al., 2017; Rendon et al., 2012). Only two studies provided a clear strategy for the withdrawal of prompts through decreasing assistance gradually until mastery was achieved (Jackson et al, 2014; Montoya-Rodriguez & Cobos, 2016), highlighting that the training strategies in the majority of studies were probably not as effective as they might have been at moving from the control of the prompts to the natural discriminative stimuli for the response.

In terms of considering how well the studies promoted generalisation of acquired skills to other settings (including everyday life) and the maintenance of those skills, generalisation can be promoted by measures such as the use of multiple examples

43

of the material, multiple trainers and multiple learning contexts (including everyday life contexts) (Stokes & Baer, 1977). Maintenance can be promoted by ensuring that the skill is prompted in everyday contexts and that exercising the skill is reinforced in everyday contexts (Stokes & Baer, 1977). The studies reviewed here took few steps to promote generalisation and maintenance. Multiple exemplars were utilised in only one study (Lovett & Rehfeldt, 2014), with six studies including a generalisation probe, post intervention. The teaching of the protocol took place in the participant's everyday environment (home or classroom) in seven of the studies (Belisle et al., 2016; Gilroy et al., 2015; Heagle & Rehfeldt, 2006; Jackson et al., 2014; Melendez et al., 2010; Montoya-Rodriguez & Cobos, 2016; Weil et al., 2011), but no study involved the training of deictic skills where they could naturally arise within this environment or any other steps to promote generalisation and maintenance.

This evaluation suggested few measures have been taken to ensure that learning the skills was engaging and reinforcing, or that prompting was used to enhance the effectiveness of learning. Very few steps were taken to enhance the generalisation of the skills to other setting (particularly everyday life) and their maintenance.

3. Is there evidence that teaching or prompting the use of deictic skills can generalise to real-life contexts, be maintained over time and have an impact on more complex behaviours?

Only one low quality study assessed whether improvements were maintained at follow-up, with deictic performance still remaining "several months" after intervention (Montoya-Rodriguez et al., 2017) with this being a significant gap in the literature.

Seven studies assessed if the skills learnt through training were generalisable to a different context when completing post-tests. One study used a different experimenter (Gilroy et al., 2015), four studies used different questions (Belisle et al., 2016; Montoya-Rodriguez & Cobos, 2016; Rendon et al., 2012; Weil et al., 2011) and one changed the context of how questions were administered (Lovett & Rehfeldt, 2014). Generalisation of perspective taking skills to a different context was reported to be obtained in five of these studies (Belisle et al., 2016; Gilroy et al., 2015; Heagle & Rehfeldt, 2006; O'Neill & Weil, 2014; Weil et al., 2011). These studies were limited in their scope of evaluating generalisation, with the majority including very similar questions, to assess improvement. The assessment of skills being used in a more naturalistic setting was included in only one study, with scenarios being used as generalisation probes throughout training (Lovett & Rehfeldt, 2014). This involved a video presentation of three scenarios that university students enacted, with each scenario requiring two questions to be answered on simple, reverse and double reverse relations. Both simple and reverse relations were reported to reach mastery with double-reverse improving across participants but not achieving the mastery criteria. The findings highlight that there were few assessments of generalisation to real-life contexts or of their maintenance.

Six studies looked at whether training of deictic relations using the Barnes-Holmes protocol had an impact on more complex behaviours using TOM tasks (Jackson et al., 2014; Lovett & Rehfeldt, 2014; Montoya-Rodriguez & Cobos, 2016; O'Neill & Weil, 2014; Rendon,et al., 2012; Weil et al., 2011).

Two studies (both moderate quality; O'Neil & Weil, 2014; Weil et al., 2011) reported an improvement of TOM tasks following mastery of deictic relational training

administered using verbal presentation and responses. Improvement on the TOM tasks were reported for typically developing young children in relation to enhanced understanding of seeing leads to knowing, predicting actions on the basis on an individual's knowledge and predicting actions on an individual's false belief (ToM; levels 3, 4 and 5) and adults diagnosed with schizophrenia and mild to moderate learning disability (TOM; levels 4 and 5) following deictic training. Most improvements of ToM tasks were found following the mastery of double reversal training. Participants on average completed the TOM tasks on three occasions during the study.

In contrast, four studies (2 low, 2 moderate quality) reported no improvement in TOM tasks following mastery of deictic relational responding. (Jackson et al., 2014; Lovett & Rehfeldt, 2014; Montoya-Rodriguez & Cobos, 2016; Rendon et al., 2012). One study included young children diagnosed with autism (Jackson et al., 2014), one with young adults diagnosed with asperger's syndrome (Lovett & Rehfeldt, 2014) and two further studies included typically developing young children with no known disabilities (Montoya-Rodriguez & Cobos, 2016; Rendon et al., 2012).

Only one study (moderate quality) assessed ToM improvement using percentages of correct responses across trials and reported no significant improvement following deictic training (Jackson et al., 2014). Furthermore, to date, no study has assessed the maintenance of such ToM perspective taking skills, and only one study has assessed if these ToM skills generalise to everyday life (Lovett & Rehfeldt, 2014). The preliminary findings would suggest that evidence of transfer of training on deictic responding to ToM tasks is mixed.

Five studies used deictic practice to enhance the distinction between self and content of thoughts as part of cognitive defusion (3 moderate, 2 high quality) (Foody et al., 2015; Foody et al., 2013; Gil-Luciano et al., 2017; Luciano et al., 2011; Ruiz & Perete, 2015). These studies prompted use of deictic skills in distinguishing between the self and experience, encouraging the move towards self-as-context and away from self-as-content. The deictic relation interventions enabled participants to observe their behaviour as something that happened THERE whilst positioning themselves as HERE. Additionally, hierarchical relations enabled participants to observe such behaviour and to differentiate this from the self (i.e. "I contain the behaviour, it is only part of me"). In three studies that utilised a non-clinical sample, significant improvements were reported following the use of deictic framing in an increased tolerance to experimentally induced pain (Gil-Luciano et al., 2017) and a significant reduction in experimentally induced stress, although this did not continue to improve following a short practice period (Foody et al., 2015). Additionally, significant reductions in believability and attempts to distract from self-criticism were recorded (Foody et al., 2013). Finally, a significant reduction in the anger episodes of a five-year-old boy were found when incorporating both deictic and hierarchical techniques combined and these were maintained at one year follow-up (Ruiz & Perete, 2015). However, the intervention was implemented by the participant's mother, included a token economy and only included one participant. As no experimental control was included, changes and the maintenance of improvement may have been due to other confounding variables.

The preliminary findings would suggest that a firm conclusion about the effectiveness of prompting the use of deictic skills in the context of highlighting the difference

between the self and experience is unable to be established. The evidence about whether the maintenance of prompting such skills is mixed and therefore the long-term benefits is unable to be determined.

DISCUSSION

Is Barnes-Holmes protocol effective in improving deictic relation performance on the task used for the training?

Few clear conclusions can be drawn in answer to this question, because of the poor quality of the evidence. Single-case experimental designs were used, with limited baseline data points and limited statistical analysis and a reliance on visual inspection alone to infer that the protocol was effective. Although the studies reported improvements, these limitations weaken confidence that these improvements can be attributed to the training.

Do the teaching strategies used in the studies follow best practice in terms of (a) improving performance on the training task (b) improving generalisation to other settings (including everyday life) and (c) maintenance over time?

Few studies used teaching strategies that followed best practice, by ensuring that learning the skills was engaging and reinforcing. Prompting was also not used effectively. The failure to address these issues within the training, may have been a factor in why a considerable number of trials were required in some studies to reach mastery, with participants losing motivation to complete the task. Few studies took any steps to promote generalisation and maintenance of the acquired skill. Evidence about the transfer of skills to more complex behaviours such as the ToM tasks was mixed. The inconsistencies across studies might be explained when taking the methodologies into account. For example, a variety of populations were included

(typically developing children, children diagnosed with autism/aspergers, children placed in institutionalised settings and adults diagnosed with schizophrenia and learning disabilities) making it difficult to compare improvements consistently across studies. Furthermore, only one study attempted to assess improvement of ToM following training and this study used basic analysis of using percentages of correct responses across trials (Jackson et al., 2014), reporting no improvement. The ToM measures utilised across studies varied and were administered over different time intervals, with some studies allowing small amounts of time to detect change in participants. This may have reduced the reliability of the measure to detect changes in ToM skills. Studies that reported an improvement in ToM performance may also have been confounded by the practice effects. For example, in the study by Weil et al., (2011), participants completed the ToM task several times, and this may have been responsible for the improvement they showed. Supporting this explanation, Montoya-Rodriguez and Cobos (2016) reported an improvement of ToM results in the control group, following repeated exposure to the task. A further potential confounding factor is that Weil and colleagues (2011) used typically developing children from an age range when ToM skills typically would develop naturally, which threatens the internal validity of the study due to possible maturation effects. Overall, one can conclude that the studies did not follow best practice in their design of the teaching protocol.

Is there evidence that teaching or prompting the use of deictic skills can generalise to real-life contexts, be maintained over time, and have an impact on more complex behaviours?

Assessment of generalisation and maintenance of the studies was very limited. The generalisation assessments lacked comprehensiveness (e.g. the assessment was confined in one study to whether the skill generalised to a different trainer) and only one study (which was of poor quality) conducted a maintenance assessment. Evidence about generalisation was mixed. Until further evidence is gathered, no clear conclusions can be drawn.

As well as an impact on ToM tasks, significant improvements were reported for increased pain tolerance (Gil-Luciano et al., 2017), reduced distress (Foody et al., 2015), anger (Ruiz & Perete, 2015), believability, and attempts to distract from self-criticism (Foody et al., 2013) following deictic framing. However, due to the small number of studies exploring a range of different factors and the lack of validity of these studies (representative population not reported) ascertaining if deictic skills can generalise to more complex behaviours of this nature cannot be determined.

Even if deictic training can have an impact on these more complex behaviours, it is not clear whether the impact is sustained. Inconsistency regarding the long-term benefits of deictic framing were found, with only Ruiz and Perete (2015) reporting improvement remaining at follow-up. This study included both deictic and hierarchical framing combined as an intervention for one participant, making it difficult to determine the impact of the deictic framing only. Improvements were not found to remain after a six-day practice period when using deictic framing alone (Foody et al., 2015). The maintenance of deictic framing requires more research to establish if there are long term benefits.

RECOMMENDATIONS

Provide more robust evidence for the effectiveness of the Barnes-Holmes protocol in teaching deictic skills.

Randomised control trials would enable selection bias to be reduced and help to determine with more confidence that the positive effects of deictic relational training are due to the intervention. The methodology could also be improved by providing detailed descriptions of the baseline characteristic of participants, thereby increasing external validity; utilising larger samples, matching participants by age and deictic performance at baseline, and using statistical analysis to assess change between experimental groups.

Improve the teaching protocol.

A variety of teaching strategies were included across studies, but a lack of consistency with best practice was found. Assessing task engagement and whether teaching materials are of interest to participants, and taking steps to enhance these where there is a problem, could improve the teaching strategy of the protocol and make training more effective. Breaking down the number of trials, alternating scenarios and/or context of training and including prompts (visual aids) could increase motivation.

Ensuring that a preference assessment is completed prior to training and including a range of artificial reinforcement, would ensure that the consequences are more

reinforcing. Including scenarios and materials tailored to the participant's interest, incorporating interactive aspects to the intervention, providing regular breaks and ensuring sessions are short could help to reduce boredom levels, frustration and keep participants on task.

Providing prompts to increase the probability of a correct response, providing immediate reinforcement following a successful response and providing clear and immediate feedback would enhance the learning process. Specifying a strategy for how to fade prompts when performance improves, would enhance the transfer from the control of the prompts to the natural discriminative stimuli for the response.

Take steps to promote generalisation and maintenance.

Generalisation can be promoted by measures such as the use of multiple examples of the material, multiple trainers and multiple learning contexts (Stokes & Baer, 1977). Generalisation to everyday life is best promoted by teaching the skills in everyday life when opportunities for using the skill occur naturally and are reinforced naturally (Stokes & Baer, 1977). Maintenance can be promoted by ensuring that the skill is prompted in everyday contexts and that exercising the skill is reinforced in everyday contexts (Stokes & Baer, 1977).

Provide more evidence about generalisation and maintenance.

More evidence is needed about the extent to which the training of these skills does generalise across contexts and is maintained over time. Assessments of

generalisation needs to be comprehensive, covering differences in persons involved, tasks and context; and maintenance needs to be assessed at several time points, controlling for maturation effects by comparing the results with matched participants who do not receive the training.

Provide more evidence about the transfer of the skills and design protocols to enhance transfer.

Future research should attempt to fill a current gap in the literature by investigating whether teaching perspective taking skills or prompting the use of existing skills can result in transfer to more complex behaviours to which deictic skills are assumed to contribute (e.g. ToM tasks, empathy, cognitive defusion). Furthermore, work is needed to devise and evaluate protocols that facilitate the transfer of skills to these more complex tasks. For example, how do we ensure that someone who is learning deictic skills transfers this to situations requiring empathy and taking the perspective of the other person? One would imagine that prompting might facilitate transfer (e.g. pointing out to the person the relevance of the deictic skills and prompting them to use them). However, these are issues that require further research.

CONCLUSION

Deficits in perspective taking skills have been associated with interpersonal communication, social cognition and poor functioning in social situations. Additionally, they may play a role in some clinical disorders and influence an individual's sense of self. It is important for future research to try and provide more robust evidence about whether these skills can be taught and whether these acquired skills can then generalise into everyday life and transfer to other more complex skills.

REFERENCES

Barnes-Homes, Y. (2001). Analysing relational frames: Studying Language and cognition in young children (Unpublished doctoral thesis: National University of Ireland, Maynooth)

Barnes-Holmes, Y., Barnes-Holmes, D., & Cullinan, V. (2001). Education. In S. C. Hayes, D. Barnes-Holmes, & B. T. Roche (Eds.) *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 181-195). New York: Plenum

Baron-Cohen, S. (2005). Autism and the origins of social neuroscience. In A. Easton & N. Emery (Eds.), *The cognitive neuroscience of social behaviour* (pp. 239–255). New York, NY: Psychology Press

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D.J. (Eds.) (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed.). Oxford, UK: Oxford University Press

Belisle, J., Dixon, M., Stanley, C., Munoz, B., & Daar, J. (2016). Teaching foundational perspective-taking skills to children with autism using the PEAK-T curriculum: single-reversal "I-You" deictic frames. *Journal of Applied Behavior Analysis,* 49(4), 965-969. http://dx.doi.org/10.1002/jaba.324

Bora, E., Eryavuz, A., Kayahan, B., Sungu, G., & Veznedaroglu, B. (2006). Social functioning, theory of mind and neurocognition in outpatients with schizophrenia; mental state decoding may be a better predictor of social

functioning than mental state reasoning. *Psychiatry Resea*rch, 145(2-3), 95-103. http://dx.doi.org/10.1016/j.psychres.2005.11.003

Chin, H.Y. & Bernard-Opitz, V. (2000). Teaching conversational skills to children with autism: effect on the development of a theory of mind. *Journal of Autism & Developmental Disorders*, 30(6):569-583

Crosbie, J. (1993). Interrupted time-series analysis with brief single-subject data. *Journal of Consulting and Clinical Psychology*, 61(6), 966-974. http://dx.doi.org/10.1037/0022-006x.61.6.966

Davlin, N., Anne Rehfeldt, R., & Lovett, S. (2011). A Relational Frame Theory Approach to Understanding Perspective-Taking Using Children's Stories in Typically Developing Children. European. *Journal of Behavior Analysis*, 12(2), 403-430. http://dx.doi.org/10.1080/15021149.2011.11434392

Downs, A., & Smith, T. (2004). Emotional Understanding, Cooperation, and Social Behavior in High-Functioning Children with Autism. *Journal of Autism and Developmental Disorders*, 34(6), 625-635. http://dx.doi.org/10.1007/s10803-004-5284-0

Foody, M., Barnes-Holmes, Y., Barnes-Holmes, B., & Luciano, C. (2013). An Empirical Investigation of Hierarchical versus Distinction Relations in a Self-based ACT Exercise, *International Journal of Psychology & Psychological Therapy*, 13, 3, 373-388

Foody, M., Barnes-Holmes, Y., Barnes-Holmes, D., Rai, L., & Luciano, C. (2015). An Empirical Investigation of the Role of Self, Hierarchy, and Distinction in a

Common Act Exercise. *The Psychological Record*, 65(2), 231-243. http://dx.doi.org/10.1007/s40732-014-0103-2

Gil-Luciano, B., Ruiz, F., Valdivia-Salas, S., & Suárez-Falcón, J. (2017). Promoting Psychological Flexibility on Tolerance Tasks: Framing Behavior through Deictic/Hierarchical Relations and Specifying Augmental Functions. *The Psychological Record*, 67(1), 1-9. http://dx.doi.org/10.1007/s40732-016-0200-5

Gilroy, S., Lorah, E., Dodge, J., & Fiorello, C. (2015). Establishing deictic repertoires in autism. *Research in Autism Spectrum Disorders*, 19, 82-92. http://dx.doi.org/10.1016/j.rasd.2015.04.004

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York, NY: Plenum Press.

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

Heagle, A., & Rehfeldt, R. (2006). Teaching perspective-taking skills to typically developing children through derived relational responding. *Journal of Early and Intensive Behavior Intervention*, 3(1), 1-34. http://dx.doi.org/10.1037/h0100321

Hendriks, A., Barnes-Holmes, Y., McEnteggart, C., De Mey, H., Janssen, G., & Egger, J. (2016). Understanding and Remediating Social-Cognitive

Dysfunctions in Patients with Serious Mental Illness Using Relational Frame Theory. *Frontiers in Psychology*, 7. http://dx.doi.org/10.3389/fpsyg.2016.00143

Higgins, J.T. & Green, S., (Eds.) (2006). *Cochrane Handbook for Systematic Reviews of Interventions 4.2.6.* In *The Cochrane Library*, Issue 4. Chichester, UK: John Wiley & Sons, Ltd.

Howlin, P., Baron-Cohen, S., & Hadwin, J. (1999). *Teaching children with autism to mind-read*. Chichester: J. Wiley & Sons.

Jackson, M., Mendoza, D., & Adams, A. (2014). Teaching a Deictic Relational Repertoire to Children with Autism. *The Psychological Record,* 64(4), 791-802. http://dx.doi.org/10.1007/s40732-014-0078

Langdon, R., Michie, P., Ward, P., McConaghy, N., Catts, S., & Coltheart, M. (1997). Defective Self and/or Other Mentalising in Schizophrenia: A Cognitive Neuropsychological Approach. *Cognitive Neuropsychiatry*, 2(3), 167-193. http://dx.doi.org/10.1080/135468097396324

Lovett, S., & Rehfeldt, R. (2014). An evaluation of multiple exemplar instruction to teach perspective-taking skills to adolescents with Asperger Syndrome. *Behavioral Development Bulletin*, 19(2), 22-36. http://dx.doi.org/10.1037/h0100575

Luciano, C., Ruiz, F., Torres, R., Martin, V., Martínez, O., & López, J. (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, 165- 182

Martin, G., & Pear, J. (1992). *Behavior modification: What it is and how to do it* (4th ed.). Engelwood Cliffs, NJ: Prentice-Hall

McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004). Perspective-Taking as Relational Responding: A Developmental Profile. *The Psychological Record*, 54(1), 115-144. http://dx.doi.org/10.1007/bf03395465

Mchugh, L., Barnes-Holmes, Y., Barnes-Holmes, D., Whelan, R., & Stewart, I. (2007). Knowing Me, Knowing You: Deictic Complexity in False belief Understanding. *The Psychological Record*, 57(4), 533-542. http://dx.doi.org/10.1007/bf03395593

Meléndez, G. (2010). Assessment and training of perspective taking: A functional analytical proposal. *Revista Diversitas–Perspectivas en Psicología*, 6(1), 141–151

Miltenberger, R.M. (2000). *Behavior modification: Principles and procedures* (2nd ed.). Belmont, CA: Wadsworth.

Montoya-Rodríguez, M., & Molina Cobos, F. (2016). Relationship between Deictic Relational Responding and Theory of Mind Tasks in Children: A Pilot Study. *The Psychological Record*, 66(4), 573-587. http://dx.doi.org/10.1007/s40732-016-0193-0

Montoya-Rodríguez, M., McHugh, L., & Cobos, F. (2017). Teaching perspective-taking skills to an adult with Down syndrome: A case study. *Journal of Contextual Behavioral Scienc*e, 6(3), 293-297. http://dx.doi.org/10.1016/j.jcbs.2017.04.012

Montoya-Rodríguez, M., Molina, F., & McHugh, L. (2016). A Review of Relational

Frame Theory Research into Deictic Relational Responding. *The

Psychological Record,* 67(4), 569-579. http://dx.doi.org/10.1007/s40732-016-

0216-x

Morse, T.E., & Schuster, J.W. (2004). Simultaneous prompting: A review of the

literature. *Education and Training in Mental Retardation*, 39, 153-168.

National Institute for Health and Care Excellence (2012)*. Methods for the

development of NICE public health guidance (third edition) (PMG4).* Retrieved

from https://www.nice.org.uk/process/pmg4/resources/methods-for-the-

development-of-nice-public-health-guidance-third-edition-pdf-2007967445701

O'Neill, J., & Weil, T. (2014). Training Deictic Relational Responding in People

Diagnosed with Schizophrenia. *The Psychological Record,* 64(2), 301-310.

http://dx.doi.org/10.1007/s40732-014-0005-3

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?

Behavioral and Brain Sciences, 1(04), 515.

http://dx.doi.org/10.1017/s0140525x00076512

Rehfeldt, R. A., Barnes-Holmes, Y. (2009). *Derived relational responding:

Applications for learners with autism and other developmental disabilities*.

Oakland, CA: New Harbinger Publications, Inc.

Rehfeldt, R., Dillen, J., Ziomek, M., & Kowalchuk, R. (2007). Assessing Relational

Learning Deficits in Perspective-Taking in Children with High-Functioning

Autism Spectrum Disorder. *The Psychological Record,* 57(1), 23-47.

http://dx.doi.org/10.1007/bf03395563

Rendón, M. I., Soler, F. & Cortés, M (2012). Simple deictic relations, perspective and

social competence. *Suma Psicológica*, 19(2), 19-37. Retrieved April 25, 2018,

from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-

43812012000200002&lng=en&tlng=es.

Ruiz, F.J. & Perete, L. (2015). Application of a relational frame theory account of

psychological flexibility in young children. Psicothema, Vol 27, No. 2, 114-119.

doi: 10.7334/psicothema2014.195

Stokes, T.F., & Baer, D.M. (1977). An implicit technology of generalization. *Journal of

Applied Behavior Analysis,* 10, 349-367

Tate, R., Rosenkoetter, U., Wakim, D., Sigmundsdottir, L. Doubleday, J., Togher, L.,

Perdice, M. (2015). *The Risk-of-Bias in N-of-1 Trials (RoBiNT) Scale: An

Expanded Manual for the Critical Appraisal of Single-Case Reports*. Australia:

The PsycBITE Group.

Vilardaga, R. (2009). A Relational Frame Theory account of empathy. International

Journal of Behavioral Consultation and Therapy, 5(2), 178-184.

http://dx.doi.org/10.1037/h0100879

Villatte, M., Monestès, J., McHugh, L., Freixa i Baqué, E., & Loas, G. (2008).

Assessing deictic relational responding in social anhedonia: A functional

approach to the development of theory of mind impairments. *International

Journal of Behavioral Consultation and Therapy*, 4(4), 360-373.

http://dx.doi.org/10.1037/h0100867

Villatte, M., Monestès, J., McHugh, L., Freixa i Baqué, E., & Loas, G. (2010).

Adopting the perspective of another in belief attribution: Contribution of

Relational Frame Theory to the understanding of impairments in

schizophrenia. *Journal of Behavior Therapy and Experimental Psychiatry*,

41(2), 125-134. http://dx.doi.org/10.1016/j.jbtep.2009.11.004

Weil, T., Hayes, S., & Capurro, P. (2011). Establishing a Deictic Relational Repertoire

in Young Children. The Psychological Record, 61(3), 371-390.

http://dx.doi.org/10.1007/bf03395767

PAPER TWO

EVALUATION OF IMPLICIT SELF-REFERENTIAL THOUGHTS IN PEOPLE WITH DEPRESSION AND THEIR ASSOCIATION WITH COGNITIVE FUSION, SELF-ESTEEM AND PSYCHOLOGICAL DISTRESS

ABSTRACT

Background

Acceptance Commitment Therapy (ACT) provides interventions on increasing psychological flexibility. Within ACT, a process known as fusion, suggests that individuals attach to internal experiences and the content of their thoughts. An ACT intervention known as defusion, involves detaching oneself from the content of thoughts and observing occurrences from a distance. Limited research has assessed whether those who fuse with their thoughts, experience greater psychological distress. The current research explores this relationship to ascertain whether there is experimental evidence that could justify this therapeutic technique.

Aim

To assess whether people who fuse with their thoughts are more likely to experience psychological distress displayed through symptoms of depression, psychological inflexibility, and have difficulties with self-esteem.

A secondary aim is to assess whether the Implicit Relational Assessment Procedure (IRAP) can help to account for more variance of severity of depression than self-report measures alone.

Method

Thirty-seven participants who were accessing mental health services for symptoms of depression, completed the study. Five self-report measures and one IRAP task that

was designed to assess implicit repertoires of responding to co-ordination and containment statements were completed.

Results

Lower levels of self-esteem and psychological flexibility and greater fusion were associated with experiencing a higher level of depression as measured by the DASS-21. However, the IRAP was unable to account for more of the variance of depression than self-report measures alone.

Conclusion

By addressing the limitations highlighted, the IRAP may be enhanced to become a tool that can increase understanding about the functional processes underlying fusion.

INTRODUCTION

Acceptance Commitment Therapy (ACT; Hayes, Strosahl & Wilson, 1999) is a behaviour therapy that focuses on increasing psychological flexibility. This requires a person to not attempt to avoid or control their unwanted internal experiences, but to accept their present experience and choose behaviour that is value oriented (Foody, Barnes-Homes, Barnes-Holmes & Luciano, 2013). To achieve a more meaningful life and greater psychological flexibility, ACT utilises six processes; acceptance, cognitive defusion, self-as-context, being present, valued goals and committed action (Hayes, Pistorello & Levin, 2006).

The theoretical basis of ACT is informed by Relational Frame Theory (RFT; Hayes, Barnes-Holmes & Roche, 2001). This theory provides "a framework to help understand the relationship between language and cognition" (Duff, Larson & McHugh, 2016). RFT proposes that human suffering is formed through the development of language, and the ability that we have to articulate relationships between stimuli. These relationships have been learnt through reinforcement in our own developmental history and have an influence on behaviour (Hughes & Barnes-Homes, 2016). Language comprises of a variety of patterns in relational responses, which are defined as relational frames, for example; co-ordination (A is the same as B), opposition (A is the opposite to B), distinction (A is different to B), comparison (A is better than B), hierarchy (A is part of B) and deictic/perspective taking (which defines the relationship from the perspective of the speaker, 'I-YOU', 'HERE-THERE', and 'NOW-THEN; (Montoya-Rodriguez, Molina & McHugh, 2016). This model can

help us identify how, as humans, we relate to internal experiences, and how we develop an understanding of our sense of self. For example, relationships of co-ordination are made up of a combination of internal experience and how we perceive the world around us. For example, someone thinking 'I am useless', may perceive this to be an accurate supposition, and will therefore be more inclined to avoid specific situations, that could reinforce this view.

The attachment to thoughts and feelings in this way from an ACT perspective is called cognitive fusion. Whilst this sense of self can be useful at times, it can also assist in a sense of self that is enmeshed with the content of thoughts and feelings, known as self-as-content. This can cause the development of a negative self-image which is detached from external experiences and can, in turn, have an influence upon behaviour (Foody, Barnes-Homes & Barnes-Homes, 2012).

The broad process of attempting to move from a self-as-content perspective to develop a 'self-as-context' perspective is similar to what is known within ACT interventions as cognitive defusion (Healy et al., 2008). This involves viewing the self as a place where internal experiences are contained separately from the contents of their thoughts (i.e. I am bigger than this thought") and is referred to as an observing self. In viewing thoughts from this perspective, an individual is less likely to become 'caught up' or fused with their thinking (Harris 2006). By doing this, one is able to perceive thoughts as nothing more than words and respond to thoughts in terms of workability rather than how true they are (Hayes, Pistorello & Levin, 2006).

RFT provides an explanation of how the process of defusion and self-as-context occurs through deictic and hierarchical responding (Luciano et al., 2011). This theory

suggests that people who view their thoughts in relation to self-as-content and are fused with their thoughts, perceive that the content of such thoughts are HERE and NOW, therefore coordinating with their thoughts. In contrast, people who view their thoughts in relation to self-as-context and are defused from such thoughts are suggested to perceive the self and being HERE and NOW, but the thought and its content being THERE and THEN. This sense of self-as-context takes a hierarchical perspective in relation to internal experiences, viewing the self as greater than the content of thoughts experienced with the individual containing the thoughts. This relationship with psychological content is perceived as the most helpful way to relate to internal experiences. Therefore, ACT encourages clients to defuse from their thoughts, by viewing such thoughts in a more flexible way, with this in turn being associated with reduced psychological distress (Zettle, Rains & Haye, 2011).

It has been hypothesised that individuals who are less psychologically flexible and fuse more with their thoughts are at greater risk of a variety of mental health difficulties (Kashdan, 2010). One such mental health difficulty is depression, which is frequently characterised by less flexible thinking patterns and a greater negative bias towards the self, others, and the future (Clark, Beck & Alford, 1999). Previous research would suggest that individuals who fuse with negative thoughts, perceive such thoughts as more believable, with this being a predictor of psychological distress (Duff, Larson & McHugh, 2016). Furthermore, individuals who more frequently attach to the content of their thoughts, particularly those that are negatively associated, have been found to experience reduced wellbeing. Additionally, reduced wellbeing has been associated with individuals who perceive self-as-content thoughts, of a positive nature (Atkins & Styles, 2016). This would suggest that the

rigidity of perceiving self-referential thoughts as factual rather than the content of such thoughts, may influence wellbeing. These findings support defusion techniques that attempt to increase psychological flexibility. However, limitations of the research to date include the issue that findings are predominantly linked to non-clinical populations. Therefore, it would seem advantageous to explore this relationship with individuals experiencing clinical levels of psychological distress, such as depression, as ACT interventions are commonly used in their treatment.

Research in the field of ACT focuses on examining 'what' makes interventions, such as defusion, effective? Inductive approaches using micro-studies on each of the six ACT processes are perceived to be a way of increasing understanding to ascertain if each process works and is psychologically active, in the way that the theory would propose (Barlow, Sauer-Zavala, Carl, Bullis & Ellard, 2013). This bottom-up approach aims to use such research to build a conceptual analysis. This process is acknowledged to be challenging, with recognition that psychological processes are not easy to access. For example, the completion of self-report measures alone is argued to not capture the psychological process of fusion, but instead captures the behaviour of completing a questionnaire, therefore not measuring a functional process (Barnes-Holmes, Hussey, McEnteggart, Barnes-Holmes, & Foody, 2016). Therefore, other methods, in addition to self-report measures are required to assess ACT processes to enable more robust findings.

As ACT interventions focus on how individuals associate with their feelings and thoughts, it is important to use measures that can detect implicit thoughts, attitudes, and beliefs when assessing outcomes (Hussey & Barnes-Holmes, 2012). Whilst self-report questionnaires can measure these explicitly, they rely on an individual's ability

to recognise their own thoughts, and how such thoughts may precipitate relapse, which may not always be accessible (Vahey, Nicholson & Barnes-Holmes, 2015). Furthermore, these measures can be sensitive to social desirability bias (Power, Barnes-Holmes, Barnes-Holmes & Stewart, 2009). Research would suggest that procedures that access implicit processes, which may occur subconsciously, could reduce the susceptibility of providing socially desirable answers (Wiers, Treachman & De Houwer, 2007). The Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006) is a validated tool (r = .45, Vahey, Nicholson & Barnes-Holmes, 2015), that enables access to implicit processes that may not typically be available, due to them occurring out of awareness. This procedure is a latency-based behavioural measure that emerged directly from RFT (Hayes, Barnes-Holmes & Roche, 2001). RFT suggests language and thoughts are developed from stimulus relations. The aim of the IRAP is to measure the strength of such relations using a computer-based task. During this task, participants are required to accurately and quickly respond to trials of stimulus pairing, selecting one of two responses under time pressure. The fundamental assumption when completing the IRAP is that a quicker response to stimulus relations will occur in response to stimuli that is coherent with previous relational histories. The difference in response latencies, between the two types of stimulus relations is defined as the IRAP effect. Therefore, this tool may enhance our understanding of how individuals relate to their thoughts and increase participant response accuracy, through the reduction of social desirability bias, (Power, Barnes-Holmes, Barnes-Holmes & Stewart, 2009).

A key part of the research is to identify whether there is sufficient evidence that ACT interventions are justifiable when used to support individuals experiencing

depression. A primary aim for the present research will be to assess whether people who fuse with their thoughts are more likely to experience psychological distress displayed through symptoms of depression, psychological inflexibility, and have difficulties with their self-esteem. This will be explored in a sample of individuals currently experiencing symptoms of depression.

Both self-report questionnaires and the IRAP will be utilised to measure both implicit and explicit attitudes and beliefs. It is hypothesised that individuals who fuse more with their thoughts, will associate with co-ordinated statements. A secondary aim of the research will be to assess whether the IRAP can help to account for more variance in the severity of depression than self-report measures alone.

METHOD

Participants

Ethical approval was granted by the National Health Service Research Ethics Committee (Appendix 4). Fifty-three participants (44 White European, 5 Indian, 4 Black Caribbean) were identified and approached in the first instance by two mental health charities situated in the West Midlands using the inclusion and exclusion criteria (Appendix 5). Participants who met the criteria were provided with an invitation to participate, which included a consent form to be contacted by the researcher and a participant information document (Appendix 6 & Appendix 7). The sample consisted of 18 males, 34 females and 1 participant who did not wish to be gender assigned. The age ranged from 20 to 67 years (M = 40.82) with level of depression severity, as scored using the DASS-21, ranging from 10-42 (n = 4, 10-13 mild, n = 11, 14-20 moderate, n = 14, 21-27 severe and n = 24, 28+ extremely severe). All participants were accessing mental health services for psychological support for depression. Sixteen participants were excluded due to not meeting performance criteria or time constraints on the IRAP (30%, see below for further description). Therefore, data from 37 participants (12 males and 24 females and 1 non-gender assigned) were analysed.

An independent t-test was used to assess for differences between completers (i.e. those meeting the criteria for valid interpretation of the IRAP) and non-completers of the IRAP. Non-completers were found to be significantly older, report greater levels

of stress, anxiety, psychological inflexibility and total level of distress compared to

completers (see Table 10). No differences were found between genders or ethnicity.

Table 10

Mean Differences and Standard Error of Self-Report Measure Scores and Age of Completers and Non-Completers

		n	Mean	Std. Deviation	t	df	Sig. (2-tailed)	Mean difference
Depression	Non-Completer	16	28.82	9.62	1.733	51	.089	4.38
	Completer	37	24.44	8.08				
Stress	Non-Completer	16	29.29	8.00	2.042	51	.046*	4.52
	Completer	37	24.78	7.29				
Anxiety	Non-Completer	16	21.41	9.79	2.39	51	.020*	6.08
	Completer	37	15.33	8.04				
Total DASS	Non-Completer	16	79.53	24.32	2.52	51	.015*	14.97
	Completer	37	64.56	18.04				
SES	Non-Completer	16	12.35	4.80	1.33	50	.188	1.78
	Completer	37	10.57	4.38				
CFQ	Non-Completer	16	38.41	11.61	.059	51	.953	.148
	Completer	37	38.26	6.60				
AAQ	Non-Completer	16	39.41	7.76	2.31	51	.025*	4.80
	Completer	37	34.61	6.71				
SDS	Non-Completer	16	18.47	4.11	2.84	49	.007*	3.47
	Completer	37	15.00	4.11				
Age	Non-Completer	16	47.76	13.42	.368	51	.006*`	10.24
	Completer	37	37.53	11.63				

Key:

DASS = Depression, Anxiety and Stress Scale; SES = Self-Esteem Scale; CFQ = Cognitive Fusion Questionnaire; AAQ = Acceptance and Action Questionnaire; SDS = Social Desirability Scale

*Significant at 0.05

Setting

All participants completed the experiment individually, in a quiet room with the experimenter present for the duration of the sessions. The participants were seated at a table with a laptop positioned in front of them. Sessions ranged from 45-60 minutes to complete.

Apparatus and Materials

The IRAP 2009 program (Barnes-Holmes) included stimuli that contained two elements. The first being the scope of the statement which could be either universal (*I really am, truly am, definitely am*) (see Figure 2a and 2b) or partial (*I only think I am, I just feel I am, I merely suspect*) (see Figure 2c and 2d), and the second element being the hedonic tone which could be either positive (see Figure 2a and 2d) or negative (see Figure 2b and 2c). Through response to this stimuli, participants expressed whether they co-ordinated, representing greater fusion, with their thoughts (all-positive or all-negative), by way of over identification with a particular construct or psychological characteristic (see Figure 2a and 2b) or expressed containment, representing a self-as-context stance (part-positive or part-negative), by recognising that a particular construct or psychological characteristic might be held or expressed in a particular situation or context and less so or not at all in other situations and contexts (see Figure 2c and 2d). Both co-ordinated and containment stimuli were presented with six positive adjectives (*strong, good, nice, attractive, brave, useful)* and six negative adjectives (*bad, weak, nasty, ugly, cowardly, useless)* and two response options (*True* and *False).* The IRAP composed of four trial types; all-positive, all-negative, part-positive and part-negative (see Figure 2). The IRAP

program presented all stimuli and recorded all responses including latency and
accuracy.

2a, Universal and Positive
I truly am **Strong**
Select 'd' for   Select 'k' for **True**   **False**

2b, Universal and Negative
I really am **Bad**
Select 'd' for   Select 'k' for **True**   **False**

2c, Partial and Negative
I merely suspect that I am **Weak**
Select 'd' for   Select 'k' for **True**   **False**

2d, Partial and Positive
I only think I am **Useful**
Select 'd' for   Select 'k' for **True**   **False**

Figure 2

An Example of Four IRAP Trials

The Top Illustrations are Co-ordination Statements and the Bottom Illustrations are
Containment Statements

The Self-Esteem Scale (SES) (Rosenberg, 1965). This self-report scale comprised
of ten items that assessed self-worth and esteem. This was measured using a 4-
point Likert scale ranging from strongly agree to strongly disagree. High scores
indicated higher levels of self-esteem. Internal consistency for this measure was
reported at $a = .77$ (Rosenberg, 1965).

The Cognitive Fusion Questionnaire (CFQ) (Gillanders et al., 2014) comprised of
seven questions assessing to what extent an individual fused with their thoughts.

This was measured on a 7-point Likert scale ranging from never true to always true. Higher scores indicated greater fusion with thoughts and lower psychological flexibility. Internal consistency reliability was been reported at *a* =.87 (McCracken, DaSilva, Skillicorn & Doherty, 2014).

The Marlowe Crowne Social Desirability Scale (MC-SDS) (Crowne & Marlowe, 1960) comprised of 33 items, which assessed whether participants were answering questions truthfully or misrepresenting themselves to protect their self-presentation. Responses to this self-report measure were answered yes or no, with test re-test reliability reported to be 0.79 (Beretvas, Meyer & Leite, 2002).

Depression, Anxiety and Stress Scale (DASS-21) (Henry & Crawford, 2005) comprised of 21-items measuring three aspects; negative emotional states of depression, anxiety and distress and overall psychological distress. This was measured using a 4-point Likert scale ranging from never to almost always. Higher scores indicated a greater level of severity of each measure. The Depression subscale had a sensitivity of 57.0% and a specificity of 67.0%. The Anxiety subscale had a sensitivity of 86.0% and a specificity of 64.0% (Mitchell, Burns & Dorstyn, 2008).

The Acceptance and Action Questionnaire (AAQ-II) (Bond et al., 2011) was a 7-item measure, assessing psychological inflexibility and experiential avoidance. This was measured using a 7-point Likert scale ranging from never true to always true.

Higher scores indicated greater psychological inflexibility. This measure reported

internal consistency of *a* =.84 and test-retest reliability .81 (Bond et al., 2011).

Procedure

Those individuals who expressed an interest were contacted by the researcher to

discuss the research and arrange an appointment. Consent to participate was

obtained prior to completion of the questionnaires (Appendix 8). Prior to the practice

block, participants were verbally instructed on how to complete an IRAP.

Participants were provided with visual examples of what the screen would look like

(see Figure 2) and how to respond to each statement by selecting key *D* for true or

key *K* for false, as appropriate, to the relation condition. Four examples were worked

through with the participants prior to commencing the computer procedure, with an

explanation provided if the response was incorrect. Participants were informed that

the tasks would alternate between answering a block containing trials from the

perspective of confirming very high positive self (e.g. "I really am strong") with fleeting

self-negative relations (e.g. I merely suspect that I am weak), named block A,

followed by a block of confirming very high negative self (e.g. "I truly am weak") with

fleeting self-positive relations (e.g. "I only think I am strong"), named block B. For

example, participants responding to a trial by asking them to answer as if they had

very high positive self with fleeting self-negative relations, would respond as follows:

all-positive statement/true, all-negative statement/false, part-positive statement/false

and part-negative statement/true. The IRAP attempted to assess the strength of

participant's relationships with these trial types, by presenting alternate blocks,

requiring opposing answers, with the stance that participants would respond quicker

to statements that were in coherence with their learning history. An incorrect

response resulted in a red X appearing on the screen, which remained on screen until the correct response was provided. A correct response resulted in the screen moving to the next trial. The participants were made aware of the criterion for fast (</= 4000ms) and accurate (>/= 70%) responses to each specific block. The practice block initially included two blocks of 24 trials but if either of the criteria were not met, the participants were provided with feedback and the opportunity to respond to a further four practice blocks to achieve the accuracy and response latency criteria (up to a total of six practice blocks). If the criteria were not met, participants were thanked for their time and debriefed, with their data excluded from the data set. Participants who met criteria continued to the six test blocks, each including 24 trials. In order to progress through the test blocks, no additional performance criteria were utilised. The IRAP program provided performance feedback to participants automatically at the end of each block to encourage criteria to be maintained. Additionally, accuracy of answers and response latency at the end of each trial was automatically recorded.

IRAP data

Sixteen participants (30%) failed to meet the criteria to pass the practice block and were therefore excluded from the analyses. Of the 37 remaining, the following exclusion criteria was applied:

Test blocks were only included if a 70% accuracy response was reached. Participants who did not meet the criteria of 70% accuracy on one of the three test blocks had this block excluded from the data set. The remaining two trial blocks that met the criteria for these participants, were included in the analysis ($N = 6$). The final number of participants included in the analyses was $N = 37$.

D-Scores

To be consistent with previous IRAP studies, response latency data, defined as the time in milliseconds from the presentation of the stimuli to the initial response, were transformed into D scores for each of the four trial-types (see Barnes-Holmes, Barnes-Holmes, Stewart & Boles, 2010 for details). This analysis entailed quantifying the differences between the IRAP trial types and recording the effect of the differences between mean response latencies between A and B trial blocks using the D algorithm developed by Greenwald, Nosek & Banaji (2003).

In establishing a D score, all outliers with latencies above 10,000ms were removed and a difference score was calculated across each of the four trial types. This

required finding the difference between one block and its associated block in the pair. Conducting this process minimised confounding variables such as age, cognitive ability and motor skills, therefore reducing contamination of the data.

The basic assumption of the *D* scores was that faster responding would occur on the blocks that were more coherent with the participants learning history. Across all trials, a positive *D*-score score of greater than 0, indicated a bias in responding quicker to block A trials compared to block B (i.e. all-positive/true, all-negative/false, part-positive/false and part-negative/true) and a negative *D*-score indicated a bias in responding quicker to block B trials compared to block A (i.e. all-positive/false, all-negative/true, part-positive/true, and part-negative/false). Scores close to zero indicated that limited difference was found when responding to block A compared to block B.

Analytic Strategy

Consistent with previous research, the IRAP data was analysed according to the following procedure. Initially, one sample t-tests were conducted to establish the significance of the difference of *D*-scores between each condition. To make the interpretation of results easier to understand, trial types were inverted (multiplied by -1) to create a common axis (Hussey, Thompson, McEnteggart, Barnes-Holmes & Barnes-Holmes, 2015). A one-way ANOVA examined the effect of the 4 trial-types to determine if there was a significant effect for trial type. A correlation examined the relationship between the self-report measures and *D*- scores. Finally, hierarchical

logistic regression analysis examined if response biases on the IRAP increased

prediction of depression status beyond self-report measures alone.

IRAP

Four D-IRAP scores were calculated for each trial type (see above) with mean and

standard error for each trial-type illustrated in Figure 3. Additionally, the last two trial-

types were inverted to help make reading of the graph easier to interpret. Therefore

a positive D score, displayed as a bar going up, indicated a positive bias and this was

only seen when participants were required to coordinate phrases about a completely

positive self-image. All of the other three trial-types displayed a negative bias,

displayed as a bar going down. A D score closer to zero for each trial-type indicated

less difference between the response times on blocks A compared to blocks B, but

the larger the difference from zero, the greater the difference between the two

blocks. The largest D score of the all-positive trial displayed the greatest difference in

response times between trials A and B.

Three trials displayed a negative bias with only all-positive trial responding in a

positive bias. Four one-sample t-tests indicated trial-types 1 to 3 all differed

significantly from zero. The co-ordination with a positive self-statements trial

revealed a significant bias to respond "true" quicker than "false," $t(36) = 5.84$,

$p<0.001$. The co-ordination with a negative self-statements trial revealed a

significant bias to respond "true" quicker than "false," $t(36) = -2.58$, $p = 0.143$.

Therefore, participants co-ordinated with their sense of self with both positive and

negative statements suggesting fusion. Participants displayed a significant bias of

selecting "false" quicker than "true" when responding to containment positive self-statements, $t(36) = 3.60$, $p = 0.010$ disconfirming a sense of self that contained positive thoughts. Finally, participants displayed a bias of selecting "true" quicker than "false" when responding to containment negative self-statements, however, this was not significant $t(36) = 0.58$, $p = 0.564$.

The final two scores were inverted (multiplied by -1) to assist in easier interpretation (Hussey, Thompson, McEnteggart, Barnes-Holmes & Barnes-Holmes, 2015) such that a positive direction demonstrated a positive bias towards oneself and a negative direction demonstrated a negative bias towards oneself (see Figure 3). Only a positive bias was found when participants were responding to themselves as all-positive. A one-way repeated measures analysis of variance (ANOVA) yielded a main effect for the all-positive trial type, reporting a significant bias for selecting 'true' more quickly than 'false' when responding to co-ordination with positive self-statement trials $F(3,36) = 13.58$ $p<0.001$. Therefore, a faster than average reaction time was indicated for this response, indicating less processing time required and greater association with these statements. However, in all other trials, participants responded with a negative bias towards the self. For example, participants coordinated with the negative self-image statements and had a negative bias when responding to the containment self-statements, reporting they did not contain positive thoughts but did contain some negative thoughts.
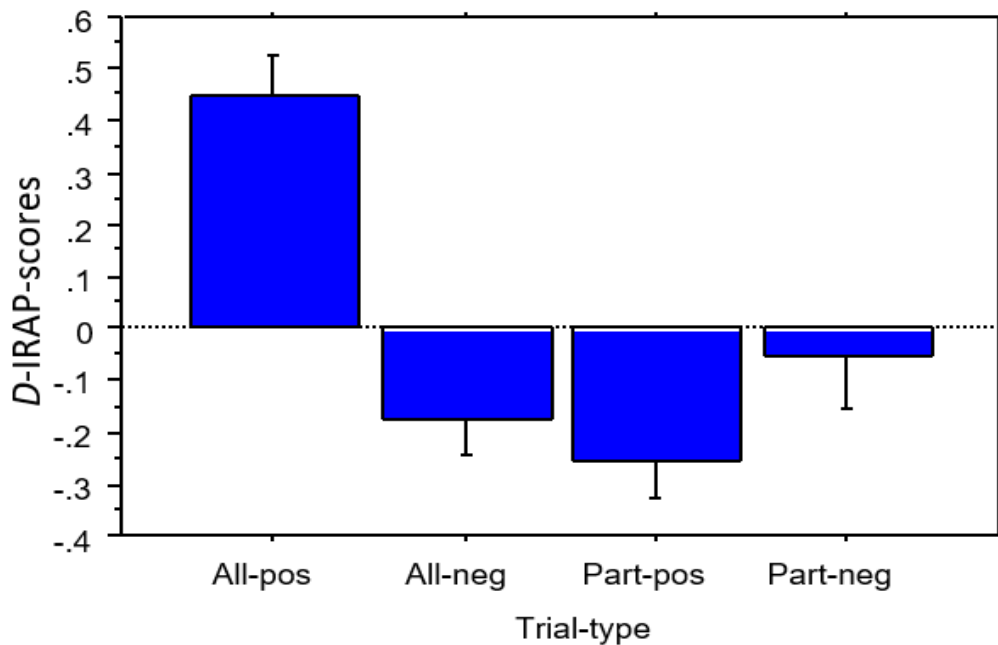
Figure 3

Mean Inverted D Scores and Standard Error of Trial-Type. Positive D Score Indicates Positive Bias

Correlations between self-report measures and the IRAP

The relationship between IRAP *D*-scores and four self-report measures was evaluated using Pearson correlation (see Table 11). A significant positive correlation involved the all-positive trial type and the SES [r .333, $p=$ 0.047] and a significant negative correlation involved the all-negative trial type and CFQ [r -.357, $p =$ 0.030], displaying predictive validity. The D-IRAP scores did not significantly correlate with the AAQ or DASS measures.

84

Table 11

Results of Correlations Between the IRAP Trial-Types and Self-Report Measures

		SES	CFQ	DASS	AAQ	All-positive	All-negative	Part-positive
CFQ	Pearson Correlation	-.403[*]						
	Sig. (2-tailed)	.015						
DASS	Pearson Correlation	-.467[**]	.334[*]					
	Sig. (2-tailed)	.004	.043					
AAQ	Pearson Correlation	-.385[*]	.721[**]	.419[**]				
	Sig. (2-tailed)	.020	.000	.010				
All-positive	Pearson Correlation	.333[*]	-.183	-.149	-.136			
	Sig. (2-tailed)	.048	.279	.379	.422			
All-negative	Pearson Correlation	.186	-.357[*]	-.082	-.218	.398[*]		
	Sig. (2-tailed)	.279	.030	.632	.194	.015		
Part-positive	Pearson Correlation	.057	-.001	.019	.234	.134	-.026	
	Sig. (2-tailed)	.743	.995	.911	.163	.428	.878	
Part-negative	Pearson Correlation	.169	-.106	-.072	-.173	.325[*]	.413[*]	-.313
	Sig. (2-tailed)	.323	.533	.671	.306	.049	.011	.059

*p<.05, **p < 0.01

85

Correlations between D-IRAP scores

Further correlations between the different trial types of *D-* scores were evaluated. Of the six relationships analysed, three were found to be significantly correlated. An increased response bias to responding to the co-ordination trial-type all-positive was positively correlated with all-negative [r .398, p = 0.014] displaying fusion for both positive and negative thoughts. Trial-type all-negative positively correlated with part-negative [r .413, p = 0.01] suggesting that the more participants perceived negative thoughts and coordinated with them, the more they rejected containment statements that part of them was negative. An unexpected finding of all-positive was correlated with part-negative [r .325, p = 0.049], suggesting trial-types interacted with each other and were not functionally independent of each other.

Correlation between self-report measures

The relationship between DASS, AAQ, CFQ and SES was evaluated using Pearson correlations. All four measures significantly correlated with each other displaying good convergence with theoretically predicted outcomes. As hypothesised, cognitive fusion in service users with depression was associated with lower self-esteem, psychological distress and inflexibility of cognitions (see Table 11).

Prediction of depression using D-IRAP scores

Four hierarchical multiple regressions were carried out to determine if *D*-scores from the four trial-types could be predictive of depression scores to a greater extent than

using the DASS21 depression subscale alone (see Table 12). This statistical method was used due to the recognition that all self-report measures correlated with each other. Each of the four self-report measures (AAQ, CFQ, SES or DASS) were individually entered into step 1 of the analysis, with trial type of the *D*-scores (all-positive, all-negative, part-positive and part-negative) entered as step 2. This analysis was used to reduce the risk of a type one error due to collinearity. *D*-scores did not significantly predict depression to a greater extent than any of the self-report measures alone (p's >.42).

Table 12

Results of Hierarchical Regression Analysis Predicting Depression Outcome using the depression subscale of the DASS 21

	Model		B	SE B	β	Adjusted R^2
SES	1	SES	-1.00	0.27	-0.54**	.287*
	2	SES	-1.00	0.30	-0.54*	.014[a]
		all positive	-0.45	3.17	-0.03	
		all negative	-1.63	3.41	-0.09	
		part positive	0.76	3.08	0.04	
		part negative	1.77	2.49	0.13	
AAQ	1	AAQ	0.50	0.19	0.41*	.171*
	2	AAQ	0.51	0.20	0.42*	.036[a]
		all positive	-2.30	3.14	-0.14	
		all negative	-0.68	3.61	-0.04	
		part positive	-1.58	3.31	-0.09	
		part negative	1.25	2.54	0.10	
CFQ	1	CFQ	0.48	0.19	0.40*	.156*
	2	CFQ	0.47	0.21	0.39*	.020[a]
		all positive	-2.61	3.19	-0.15	
		all negative	0.55	3.83	0.03	
		part positive	0.20	3.29	0.01	
		part negative	0.95	2.60	0.07	
DASS	1	DASS	0.35	0.05	0.78**	.607*
	2	DASS	0.34	0.05	0.77**	.017[a]
		all positive	-1.16	2.17	-0.07	
		all negative	-1.80	2.45	-0.09	
		part positive	-0.16	2.22	-0.01	
		part negative	1.36	1.75	0.10	

*$p < .05$

**$p < .01$

a = R^2 change

DISCUSSION

The aim of the present study was to assess whether people who fused with their thoughts were more likely to experience psychological distress displayed through symptoms of depression, psychological inflexibility and have difficulties with their self-esteem. Additionally, the IRAP was included to assess if this procedure could account for more variance of depression, than self-report measures alone. This was explored using a sample of individuals who were currently experiencing symptoms of depression and accessing mental health services for support. As hypothesised, lower levels of self-esteem, psychological flexibility and greater fusion were associated with the experience of a greater level of depression, as measured by the DASS-21. However, the IRAP was unable to account for any additional variance of depression, than self-report measures alone.

Higher levels of cognitive fusion were found to significantly correlate with lower self-esteem and higher levels of both psychological distress and inflexibility, when measured by self-report measures. These findings would support Duff, Larson & McHugh (2016), who suggested that individuals who fuse with negative thoughts experienced higher levels of psychological distress. However, a criticism of using self-report measures alone, is the possibility of the association between self-esteem, fusion and psychological distress being overinflated (Tuijl, Verwoerd & Jong, 2017). The measure of the self in previous research, has highlighted the likelihood of response bias influencing the answering of questions, across different self-report measures, to a similar degree. For example, if a participant was biased in presenting

a more positive image of self, this would influence the responses across measures, with greater scores on self-esteem and lower scores on symptoms to a relative similar extent, therefore strengthening the association (Tuijl, Verwoerd & Jong, 2017). Therefore, the present study attempted to assess this potential bias, incorporating the Marlow social desirability questionnaire. This measure revealed that overall, participants scored within the average range, suggesting an average level of concern for socially desirable responses. In addition to this self-report measure, the present study assessed this relationship using implicit measures of the IRAP, as this procedure further reduced the susceptibility to participants providing a socially desirable response.

The relationships between self-esteem, fusion, and psychological distress were replicated by the IRAP. It showed an increased bias towards responding to trial type all-positive, being significantly positively correlated with greater self-esteem and responding towards trial-type all-negative displaying, a significant relationship with greater levels of fusion. This suggests predictive validity of the IRAP on measuring self-esteem and fusion. However, it should be noted that the IRAP in this study was not designed as a measure to capture self-esteem, but instead was aiming for hierarchy and distinguishing between those who endorsed co-ordination or containment relations with reference to self-related content. Nevertheless, it would be hypothesised that the IRAP did capture self-esteem when responding to all-positive and all-negative trial types, as these trials would be seen as very close to self-esteem type, responding previously recorded by other implicit measures such as the IAT (Greenwald & Farnham, 2000). Moreover, the current study would add

support to previous research suggesting that low self-esteem is associated with increased symptoms of depression (Sowislo & Orth, 2013).

The IRAP was used to assess the strength of association between block A and block B trial-types that included statements that were co-ordinated with a sense of self, ("I am bad/good") and that contained thoughts about a sense of self ("I merely suspect that I am good/bad.") Significant IRAP effects were observed for all-positive and all-negative trial-types, showing that participants co-ordinated with both positive and negative sense of self statements, suggesting cognitive fusion of both a positive and negative self-image. This would support previous qualitative research reporting reduced wellbeing associated with both positive and negative self-conceptualisations (Atkins & Styles, 2016). However, a main effect was reported for the all-positive trial type, suggesting a significant bias for selecting 'true' more quickly than 'false when responding to co-ordination with positive self-statement trials. This would be in contrast to the cognitive theory of depression, suggesting that individuals who experience symptoms of depression associate with a negative cognitive bias of thinking processes (Beck, 1967). It could be hypothesised that the slower response time to responding to the all-negative trials when compared to all-positive could be due to participants defending against these negative statements and the emotional response that this could generate. Participants disconfirmed part-positive thoughts, reporting a sense of self that did not contain thoughts about being positive. This was consistent with ACT approach which proposes that individuals who fuse with their thoughts endorse less containment statements.

The IRAP results predominantly reported a negative bias across three of the four trial-types (all-negative, part-positive, and part-negative). Additionally, participants

were found to fuse with thoughts. These findings would suggest that participant's sense of self would be more likely to become entangled with thoughts, perceiving them to be factual in their content. Within ACT, such thoughts about the self are perceived as stories that make up a sense of identity, referred to as *conceptualised self.* From this perspective, the results would suggest that participants experiencing symptoms of depression would have a bias to evaluate, describe, understand and explain their sense of self in a manner that could be unhelpful. The conceptualisation of self in this way would be seen to be disconnected from what is actually being experienced, which results in more helpful sources of self-conceptualisation being harder to relate to (Foody, Barnes-Holmes & Barnes-Holmes, 2012). This would be perceived to be a factor that could maintain the symptoms of depression.

Interestingly, when analysing the IRAP data, the more participants responded as all-negative, the more they rejected part-negative. Furthermore, the more responses of all-positive, the greater they confirmed part-negative. These findings were not consistent with the research hypothesis and indicate that the trial-types interacted with each other and were therefore not functionally independent of each other. An explanation of why these results may have occurred could have been due to the language used for the part-negative and part-positive trial-types and the statements for these trials being longer in length, resulting in internal bias. According to relational frame theory, a history of stimuli that is encountered more frequently in the natural environment, can influence the tendency to orient towards that stimuli to a greater extent. Co-ordination relational frames are at a lower level of relational complexity when compared to containment/hierarchical relations and are the foundations of further relational frames (Blackledge, 2003). This would suggest that

the coordination trials would have a stronger orienting function response and may be responded to more quickly, reflecting a learning history of contact with those lower level stimuli (Finn, Barnes-Holmes & McEnteggart, 2016). For example, a participants' learning history would be expected to have included higher frequency of statements such as "you are useful," when compared to "I merely suspect that you are useful." This would suggest that the containment trial types would be more difficult to work out, compared to the co-ordination trial types as this stimulus would not cohere with past history. The result that the hierarchical responses took an increased amount of time to respond to would therefore not be seen to function at pulling groups apart relationally, but rather measure a level of uncertainty.

The hypothesis that the IRAP would be able to predict depression to a greater extent than self-report measures alone was not supported. The IRAP was not found to account for significantly more variance, suggesting that the implicit measure lacked predictive validity. However, an overinflated association of the self-report measures is likely to have influenced these results, due to having common-method variance, with these measures being often validated against each other (Tuijl, Verwoerd & Jong, 2017). Previous research using the IRAP scores to assess psychopathology, incorporating contamination related trial-types to predict obsessive-compulsive tendencies, avoidant behaviour and contamination fear, has been reported to provide predictive validity above self-report measures (Nicholson, Dempsey & Barnes-Holmes, 2014). This would suggest that the IRAP as a measure of psychopathogical constructs can be effective. When investigating why the current study may display inconsistent findings to previous research, some differing factors between studies may have influenced results. Nicholson and colleagues (2014) provided limited

written instructions and counterbalancing of trials across blocks, reducing the risk of confounding variables. Previous research has reported that providing specific instructions of how to respond to statements, as in the current study, may enhance the degree to which orienting functions of the stimuli influence participant's IRAP performance (Finn et al., 2016). Future research should include limited written instructions to reduce the influence of orienting and bias responses.

The IRAP utilised in the present study was used to explore the dynamics of hierarchical relational framing to gain a greater understanding of the underlying behavioural functional processes of individuals experiencing depressive symptoms. It should be acknowledged that whilst ACT is grounded in functional contextualism and is explicit in its association between relational frame theory and how this informs human language and cognitions (Zettle, 2016), understanding these concepts in relation to "successful working," is of most importance. Therefore, aspects such as "the three selves" and defusion techniques are viewed as tools that help to increase psychological flexibility rather than corresponding with external reality (Pepper, 1942). Whilst ACT research aims to understand the processes that increase psychological flexibility, it also acknowledges that mid-level terms such as fusion are challenging to measure (Blackledge & Drake, 2013). Whilst previous research has reported that such interventions are effective (Kishita, Muto Ohtsuki Barnes-Holmes, 2014), the scientific evidence for the idea that only one type of fusion, experienced by all, is difficult to ascertain, as fusion is not a functional process. Therefore, generating a model that measures fusion is challenging. ACT as an approach recognises this and strives to continuously build on developing and defining such concepts, through addressing limitations of research findings.

There are several limitations within the present study. All participants recruited were

accessing services that provided CBT as a therapeutic approach for symptoms of

depression. This approach encouraged participants to focus on the content of

thoughts, with an aim to change thinking through restructuring (Beck, 1967). In

contrast to CBT, the current IRAP entailed participants to take a hierarchical

approach to statements presented, in line with defusion techniques that required a

focus on the context of thoughts (Harris, 2009). It could be hypothesised that this

made the containment trial-type more challenging for participants, due to having no

previous exposure to hierarchical thinking, with this having an impact on the findings.

Research utilising the IRAP to predict behavioural outcome measures between

control and defusion conditions of individuals diagnosed with social anxiety, has been

found to be successful at assessing change (Kishita, Muto Ohtsuki Barnes-Holmes,

2014). This method would be seen to address the current study's limitation, by

allowing participants to be exposed to hierarchical thinking as an intervention,

enabling a greater understanding of this technique when completing the IRAP.

Future research incorporating IRAP outcome data following a defusion intervention,

with individuals experiencing depressive symptoms, would enable the analysis of the

functional processes of private events to be better understood.

Consistent with other previous research using the IRAP, a high attrition rate of 30%

of the sample who did not meet the IRAP inclusion criteria, was observed, perhaps

because of it being a relatively challenging task (Nicholson et al., 2014). This would

question how useful the IRAP currently is as a tool to measure mental health

difficulties, with an attrition rate of 30% of a clinical population not being accessible

when utilising this procedure. Whilst the reduction in completers of the task was not

ideal and acknowledged to have an impact on statistical power, a recent review of the IRAP research suggests that this sample size was above the recommended size for this research (Vahey, Nicholson & Barnes-Holmes, 2015). When analysing the data of individuals compared to those that met criteria, participants who were non-completers reported significantly greater stress, anxiety, social desirability, psychological inflexibility and were older in age. The current study would support previous IRAP research that has reported less accurate and slower responses by individuals who are more psychologically inflexible and fuse with their thoughts (Drake, Timko & Luoma, 2016). An explanation as to why this could have occurred could include extraneous variables, such as medication administered to those individuals experiencing greater levels of distress, influencing their reaction time on the IRAP task. An alternative hypothesis is that participants became frustrated and/or bored or experienced increased levels of anxiety when not meeting the criteria for the first test block, with this impacting on their further performance. As non-completers rated significantly higher than completers, scoring on the higher end of the average range on the social desirability scale, this may have increased their concern and therefore anxiety to complete the task correctly. It would be beneficial for future research to look at factors that influence the attrition rate of the IRAP, with an aim of being able to make it more accessible, and therefore increase clinical relevance and generalisation without compromising the implicit nature. Furthermore, it would be beneficial to include pictorial stimuli, in addition to text, to increase the level of interest of the procedure as this has been found to be a useful approach with previous IRAP research (Nicholson, Dempsey & Barnes-Holmes, 2014).

To conclude, this study is the first to explore the dynamics of fusion using coordination and containment statements and self-report measures in a clinical sample of people seeking support for depression. The primary hypothesis of the research was supported, with individuals who endorsed coordination statements reporting lower self-esteem and greater psychological distress and inflexibility of cognitions. The secondary aim was not supported, with the IRAP being unable to account for more variance of severity of depression than self-report measures alone. Regardless of the limitations highlighted in the present study, the discrepancy between these findings and previous research would suggest that it would be beneficial to conduct further examinations of the IRAP. By addressing the limitations discussed, the IRAP may be enhanced to become a tool that can increase understanding about the functional processes underlying fusion. It is only through such investigations exploring these processes that we can provide scientific evidence at a functional level that ACT interventions are targeting what they set out to achieve in increasing psychological flexibility.

REFERENCES

Atkins, P., & Styles, R. (2016). Measuring self and rules in what people say: exploring whether self-discrimination predicts long-term wellbeing. *Journal of Contextual Behavioral Science*, 5(2), 71-79. http://dx.doi.org/10.1016/j.jcbs.2016.05.001

Barlow, D., Sauer-Zavala, S., Carl, J. R., Bullis J. R. and Ellard, K. K. (2013) *The Nature, Diagnosis, and Treatment of Neuroticism: Back to the Future Clinical Psychological Science* Vol 2, Issue 3, pp. 344 - 365

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. http://dx.doi.org/10.1007/bf03395726

Barnes-Holmes, Y., Hussey, I., McEnteggart, C., Barnes-Holmes, D., Foody, M. (2016). Scientific ambition: The relationship between relational frame theory and middle-level terms in acceptance and commitment therapy. In R. Zettle, D, S.C. Hayes, D. Barnes-Holmes, A. Biglan (Eds.), *Handbook of Contextual Behavioral Science* (pp.365-382). New York, NY: Wiley-Blackwell

Beck, A. T. (1967). Cognitive – Maladaptive Thoughts. In Beck, A. T. *Depression – Clinical Experimental and Theoretical Aspects*, New York: Harper and Row

Beretvas, S., Meyers, J., & Leite, W. (2002). A Reliability Generalization Study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement,* 62(4), 570-589. http://dx.doi.org/10.1177/0013164402062004003

Blackledge, J. T. (2003). An introduction to relational frame theory: Basics and applications. *The Behavior Analyst Today, 3*(4), 421-433. http://dx.doi.org/10.1037/h0099997

Blackledge, J. T., & Drake, C. E. (2013). *Acceptance and commitment therapy: Empirical and theoretical considerations*. In S. Dymond & B. Roche (Eds.), *Advances in relational frame theory and contextual behavioral science: Research and application* (pp. 219-252). Oakland, CA: New Harbinger.

Bond, F., Hayes, S., Baer, R., Carpenter, K., Guenole, N., & Orcutt, H. et al., (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. http://dx.doi.org/10.1016/j.beth.2011.03.007

Clark, D. A., Beck, A. T., & Alford, B. A. (1999). *Scientific foundations of cognitive theory and therapy of depression*. New York: Wiley

Crowne, D., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349-354. http://dx.doi.org/10.1037/h0047358

Drake, C., Timko, C., & Luoma, J. (2016). Exploring an Implicit Measure of

    Acceptance and Experiential Avoidance of Anxiety. *The Psychological Record*,

    66(3), 463-475. http://dx.doi.org/10.1007/s40732-016-0186-z

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.

    http://dx.doi.org/10.1016/j.jcbs.2015.11.003

Finn, M., Barnes-Holmes, D., & McEnteggart, C. (2016). 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. http://dx.doi.org/10.1007/s40732-017-

    0262-z

Foody, M., Barnes-Holmes, Y., & Barnes-Holmes, D. (2012). Examining the role of

    self in acceptance and mindfulness. In L. McHugh and I. Stewart, (Eds.), *The*

    self and perspective-taking. Contributions and applications from modern

    behavioral science. Oakland, CA: New Harbinger.

Foody, M., Barnes-Holmes, Y., Barnes-Holmes, D., Rai, L., & Luciano, C. (2013). An

    Empirical Investigation of the Role of Self, Hierarchy, and Distinction in a

    Common Act Exercise. *The Psychological Record*, 65(2), 231-243.

    http://dx.doi.org/10.1007/s40732-014-0103-2

Gillanders, D., Bolderston, H., Bond, F., Dempster, M., Flaxman, P., & Campbell, L.

    et al., (2014). The Development and Initial Validation of the Cognitive Fusion

Questionnaire. *Behavior Therapy*, 45(1), 83-101.

http://dx.doi.org/10.1016/j.beth.2013.09.001

Harris, R. (2006). Embracing your demons: an overview of acceptance and

commitment therapy [online]. *Psychotherapy in Australia*, 12, (4), 70-76

Hayes, S.C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A*

Post Skinnerian account of human language and cognition. New York:

Plenum Press.

Hayes, S. C., Pistorello, J., & Levin, M. (2006). Acceptance and Commitment

Therapy as a Unified Model of Behavior Change. *The Counseling*

Psychologist, 40(7), 976-1002. http://dx.doi.org/10.1177/0011000012460836

Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and commitment*

therapy: An experiential approach to behavior change. New York, NY: Guilford

Press.

Healy, H., Barnes-Holmes, Y., Barnes-Holmes, D., Keogh, C., Luciano, C., & Wilson,

K. (2008). An Experimental Test of Cognitive Defusion Exercise: Coping With

Negative and Positive Self-Statements. *The Psychological Reco*rd, 58, 623-

640

Henry, J., & Crawford, J. (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.

http://dx.doi.org/10.1348/014466505x29657

Hughes, S., & Barnes-Holmes, D. (2016). Relational Frame Theory: The Basic Account. In S. Hayes, D. Barnes-Holmes, R. Zettle, and T. Biglan (Eds.), *Handbook of Contextual Behavioral Science*. New York: Wiley-Blackwell

Hussey, I., & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure as a Measure of Implicit Depression and the Role of Psychological Flexibility. *Cognitive and Behavioral Practice*, 19(4), 573-582. http://dx.doi.org/10.1016/j.cbpra.2012.03.002

Hussey, I., Thompson, M., McEnteggart, C., Barnes-Holmes, D. and 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. http://doi.org/10.1016/j.jcbs.2015.05.001

Kashdan, T. (2010). Psychological Flexibility as a Fundamental Aspect of Health. *Clinical Psychology Review*, 20, 7, 865-878

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. http://dx.doi.org/10.1016/j.jcbs.2013.12.001

Luciano, C., Ruiz, F., Torres, R., Martin, V., Martínez, O., & López, J. (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, 165- 182

McCracken, L., DaSilva, P., Skillicorn, B., & Doherty, R. (2014). The Cognitive Fusion Questionnaire. *The Clinical Journal of Pain*, 30(10), 894-901. http://dx.doi.org/10.1097/ajp.0000000000000047

Mitchell, M., Burns, N., & Dorstyn, D. (2008). Screening for depression and anxiety in spinal cord injury with DASS-21. *Spinal Cord*, 46(8), 547-551. http://dx.doi.org/10.1038/sj.sc.3102154

Nicholson, E., Dempsey, K., & Barnes-Holmes, D. (2014). The role of responsibility and threat appraisals in contamination fear and obsessive-compulsive tendencies at the implicit level. *Journal of Contextual Behavioral Science*, 3(1), 31-37. http://dx.doi.org/10.1016/j.jcbs.2013.11.001

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. http://dx.doi.org/10.1007/bf03395684

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Sowislo, J., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin*, 139(1), 213-240. http://dx.doi.org/10.1037/a0028931

Vahey, N., 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 Psychia*try, 48, 59-65. http://dx.doi.org/10.1016/j.jbtep.2015.01.004

Van Tuijl, L., Verwoerd, J., & de Jong, P. (2018). Influence of sad mood induction on

implicit self-esteem and its relationship with symptoms of depression and

anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*.

http://dx.doi.org/10.1016/j.jbtep.2018.02.003

Wiers, R., Teachman, B., & De Houwer, J. (2007). Implicit cognitive processes in

psychopathology: An introduction. *Journal of Behavior Therapy and*

Experimental Psychiatry, 38(2), 95-104.

http://dx.doi.org/10.1016/j.jbtep.2006.10.002

Zettle, R. (2016). The Self Acceptance and Commitment Therapy. In M. Kyrios, R.

Moulding, G. Doron, S. Bhar, M. Nedeljkovic, & M. Mikulincer (Eds.), *The Self*

in Understanding and Treating Psychological Disorders. London: Cambridge

University Press.

Zettle, R., Rains, J., & Hayes, S. (2011). Processes of Change in Acceptance and

Commitment Therapy and Cognitive Therapy for Depression: A Mediation

Reanalysis of Zettle and Rains. *Behavior Modification*, 35(3), 265-283.

http://dx.doi.org/10.1177/0145445511398344

PAPER THREE

PUBLIC DISSEMINATION

DOCUMENT

This document provides an overview of both the literature review and the empirical paper, submitted in partial fulfilment of the requirements for the Doctorate of Clinical Psychology (Clin Psy.D) at the University of Birmingham.

Literature review of the effectiveness, teaching strategies, generalisation and maintenance of training deictic relations and framing

Background

Cognitive and behavioural approaches suggest perspective-taking as an important basic skill that contributes to more complex behaviours and skills. These include being able to take account of the thoughts, feelings and motivations of others in planning one's own behaviour and taking effective part in social interaction such as humour, irony and deceit. Deficits in perspective taking skills can have a serious impact on communication and can result in difficulties in social situations. Enhancing an individual's perspective-taking skills, therefore, has the potential to be an important target in clinical practice. Recent work within the context of Relational Frame Theory (RFT) has suggested that enhancing these skills may be possible using behavioural teaching methods. Training perspective-taking skills, therefore, has the potential to be an important target in clinical practice.

From an RFT account, deictic is a word that is used to specify either identity, spatial or temporal location, from the perspective of the participant. These include the relations of I-YOU, HERE-THERE and NOW-THEN (Barnes-Holmes, Barnes-Holmes, & Cullinan, 2001). Research in perspective-taking have developed a

protocol (widely known as the *Barnes-Holmes protocol*) to investigate deictic

relational responses (Barnes-Holmes, 2001). This assesses deictic relations across

three levels of task complexity; *simple*, *reversed* and *double reversed* with an aim to

support learning of how to respond relationally (McHugh, Barnes-Holmes & Barnes-

Holmes, 2004).

Aim

The literature review aimed to evaluate the effectiveness, teaching strategies,

generalisation and maintenance of training deictic relations and framing.

Method

A systematic literature search identified published journal articles reporting on

teaching deictic skills or interventions designed to encourage the use of existing

deictic skills. Seventeen articles were included in the review. The quality of each

study's methodology, outcome measures and contribution to the evidence base were

reviewed.

Results

The review found that the majority of studies were of the moderate quality range.

The protocol was found to enable participants to reach the specified mastery criteria

following deictic training. However, the quality of the evidence across studies was

poor and so no firm conclusion regarding the effectiveness, generalisation or maintenance of skills following training could be determined.

Conclusion

The review highlighted the need for future research to provide more high-quality studies using better teaching strategies, training being provided in more natural environments and examining the long term effects of the protocol following training.

Empirical Paper: Evaluation of implicit self-referential thoughts of people with depression and their association with cognitive fusion, self-esteem and psychological distress

Background

Acceptance Commitment Therapy (ACT; Hayes, Strosahl & Wilson, 1999) suggests that how people relate to their thoughts is of greater importance than the actual content of such thoughts. This approach suggests people may view the content of their thoughts as something separate from themselves and be aware that such thoughts are continuously changing. This is linked to a more secure sense of self. Alternatively, people may become attached to the content of their thoughts, believing them to be factual and identifying with whatever the content may be. This is linked to a more compromised sense of self and is known within ACT as the process of cognitive fusion.

Previous research would suggest that people who experience depression, may get 'caught up in their thoughts,' linked to a compromised sense of self, to a greater extent than people who are not depressed (Duff, Larson & McHugh, 2016). However, currently there is limited understanding about the mechanisms underlying such ways of thinking.

Aims

The aim of the study was to find out more about the way people who currently experienced symptoms of depression, associated with the content of their thoughts and to gain a greater understanding of the mechanisms underlying such ways of thinking. A secondary aim was to assess if we could predict the severity of depression, to a greater extent than self-report measures alone, using an implicit assessment procedure known as the IRAP. This could increase knowledge of how to support individuals experiencing depression and provide evidence for interventions that attempt to help these people to detach from their thoughts.

Method

Thirty-seven participants who were accessing mental health services for symptoms of depression completed the study. Five self-report measures that measured self-esteem, cognitive fusion, psychological distress, experiential avoidance, acceptance and action and social desirability were completed. Additionally, a computer task, known as the Implicit Relational Assessment Procedure (IRAP) was completed to

assess whether participants were able to separate from themselves and the content of their thoughts or tended to become attached to this content.

Results

Lower levels of self-esteem, lower ability to think more flexibly and greater levels of identifying and attaching to the content of thoughts were associated with experiencing a greater level of depression. However, the IRAP was unable to predict the severity of depression to a greater extent than self-report measures alone.

Conclusion

The study would suggest that it would be beneficial to conduct further research that addresses the limitations highlighted of the IRAP. This could enable an increased understanding about the processes underlying how people associate and get 'caught up' with their thoughts.

REFERENCES

Barnes-Holmes, Y., Barnes-Holmes, D., & Cullinan, V. (2001). Education. In S. C. Hayes, D. Barnes-Holmes, & B. T. Roche (Eds.) *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 181-195). New York: Plenum

Barnes-Homes, Y. (2001). Analysing relational frames: Studying Language and cognition in young children (Unpublished doctoral thesis: National University of Ireland, Maynooth)

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. http://dx.doi.org/10.1016/j.jcbs.2015.11.003

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

McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004). Perspective-Taking as Relational Responding: A Developmental Profile. *The Psychological Record*, 54(1), 115-144. http://dx.doi.org/10.1007/bf03395465

Appendix 1: Internal and External Validity Categories used to Assess Quality (Tate et al., 2015)

Items in RoBiNT Scale

Internal Validity Subscale

1. Design
2. Randomisation
3. Sampling behaviour (all phases)
4. Blinding patient/therapist
5. Blinding accessors
6. Inter-rater reliability
7. Treatment adherence

External validity and intrepretation subscale

8. Baseline characteristics
9. Theraputic Setting
10. Dependent Variable (target behaviour)
11. Independent variable (intervention)
12. Raw data record
13. Data analysis
14. Replication
15. Gereralisation

Appendix 2: Quality Appraisal Checklist: Quantitative Intervention Studies

(NICE, 2012)

Section 1: Population		
1.1 Is the source population or source area well described? Was the country (e.g. developed or non-developed, type of healthcare system), setting (primary schools, community centres etc.), location (urban, rural), population demographics etc. adequately described?	++ + - NR NA	Comments:
1.2 Is the eligible population or area representative of the source population or area? Was the recruitment of individuals, clusters or areas well defined (e.g. advertisement, birth register)? Was the eligible population representative of the source? Were important groups under-represented?	++ + - NR NA	Comments:
1.3 Do the selected participants or areas represent the eligible population or area? Was the method of selection of participants from the eligible population well described? What % of selected individuals or clusters agreed to participate? Were there any sources of bias? Were the inclusion or exclusion criteria explicit and appropriate?	++ + - NR NA	Comments:
Section 2: Method of allocation to intervention (or comparison)		
2.1 Allocation to intervention (or comparison). How was selection bias minimised? Was allocation to exposure and comparison randomised? Was it truly random ++ or pseudo-randomised + (e.g. consecutive admissions)? If not randomised, was significant confounding likely (−) or not (+)? If a cross-over, was order of intervention randomised?	++ + - NR NA	Comments:
2.2 Were interventions (and comparisons) well	++	Comments:

described and appropriate? Were interventions and comparisons described in sufficient detail (i.e. enough for study to be replicated)? Was comparisons appropriate (e.g. usual practice rather than no intervention)?	+ - NR NA	
2.3 Was the allocation concealed? Could the person(s) determining allocation of participants or clusters to intervention or comparison groups have influenced the allocation? Adequate allocation concealment (++) would include centralised allocation or computerised allocation systems.	++ + - NR NA	Comments:
2.4 Were participants or investigators blind to exposure and comparison? Were participants and investigators – those delivering or assessing the intervention kept blind to intervention allocation? (Triple or double blinding score ++) If lack of blinding is likely to cause important bias, score −.	++ + - NR NA	Comments:
2.5 Was the exposure to the intervention and comparison adequate? Is reduced exposure to intervention or control related to the intervention (e.g. adverse effects leading to reduced compliance) or fidelity of implementation (e.g. reduced adherence to protocol)? Was lack of exposure sufficient to cause important bias?	++ + - NR NA	Comments:
2.6 Was contamination acceptably low? Did any in the comparison group receive the intervention or vice versa? If so, was it sufficient to cause important bias? If a cross-over trial, was there a sufficient wash-out period between interventions?	++ + - NR NA	Comments:
2.7 Were other interventions similar in both	++	Comments:

groups? Did either group receive additional interventions or have services provided in a different manner? Were the groups treated equally by researchers or other professionals? Was this sufficient to cause important bias?	+ - NR NA	
2.8 Were all participants accounted for at study conclusion? Were those lost-to-follow-up (i.e. dropped or lost pre-,during or post-intervention) acceptably low (i.e. typically <20%)? Did the proportion dropped differ by group? For example, were drop-outs related to the adverse effects of the intervention?	++ + - NR NA	Comments:
2.9 Did the setting reflect usual UK practice? Did the setting in which the intervention or comparison was delivered differ significantly from usual practice in the UK? For example, did participants receive intervention (or comparison) condition in a hospital rather than a community-based setting?	++ + - NR NA	Comments:
2.10 Did the intervention or control comparison reflect usual UK practice? Did the intervention or comparison differ significantly from usual practice in the UK? For example, did participants receive intervention (or comparison) delivered by specialists rather than GPs? Were participants monitored more closely?	++ + - NR NA	Comments:
Section 3: Outcomes		
3.1 Were outcome measures reliable? Were outcome measures subjective or objective (e.g. biochemically validated nicotine levels ++ vs self-reported smoking −)? How reliable were outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated (e.g. validated against a gold standard measure or assessed for content validity)?	++ + - NR NA	Comments:

3.2 Were all outcome measurements complete? Were all or most study participants who met the defined study outcome definitions likely to have been identified?	++ + - NR NA	Comments:
3.3 Were all important outcomes assessed? Were all important benefits and harms assessed? Was it possible to determine the overall balance of benefits and harms of the intervention versus comparison?	++ + - NR NA	Comments:
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure? (e.g. a study to assess impact on physical activity assesses gym membership – a potentially objective outcome measure – but is it a reliable predictor of physical activity?)	++ + - NR NA	Comments:
3.5 Were there similar follow-up times in exposure and comparison groups? If groups are followed for different lengths of time, then more events are likely to occur in the group followed-up for longer distorting the comparison. Analyses can be adjusted to allow for differences in length of follow-up (e.g. using person-years).	++ + - NR NA	Comments:
3.6 Was follow-up time meaningful? Was follow-up long enough to assess long-term benefits or harms? Was it too long, e.g. participants lost to follow-up?	++ + - NR NA	Comments:
Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were these adjusted? Were there any differences between groups in important confounders at baseline? If so, were these adjusted for in the analyses (e.g. multivariate analyses or stratification). Were there likely to be any residual differences of relevance?	++ + - NR NA	Comments:
4.2 Was intention to treat (ITT) analysis	++	Comments:

conducted?	+	
Were all participants (including those that dropped out or did not fully complete the intervention course) analysed in the groups (i.e. intervention or comparison) to which they were originally allocated?	- NR NA	
4.3 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 (that is, it is likely to see an effect of a given size if one exists, 80% of the time) is the conventionally accepted standard. Is a power calculation presented? If not, what is the expected effect size? Is the sample size adequate?	++ + - NR NA	Comments:
4.4 Were the estimates of effect size given or calculable? Were effect estimates (e.g. relative risks, absolute risks) given or possible to calculate?	++ + - NR NA	Comments:
4.5 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? If a cluster design, were analyses of sample size (and power), and effect size performed on clusters (and not individuals)? Were subgroup analyses pre-specified?	++ + - NR NA	Comments:
4.6 Was the precision of intervention effects given or calculable? Were they meaningful? Were confidence intervals or p values for effect estimates given or possible to calculate? Were CI's wide or were they sufficiently precise to aid decision-making? If precision is lacking, is this because the study is under-powered?	++ + - NR NA	Comments:
Section 5: Summary		
5.1 Are the study results internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design?	++ + -	Comments:
5.2 Are the findings generalisable to the source	++	Comments:

population (i.e. externally valid)? Are there sufficient details given about the study to determine if the findings are generalisable to the source population? Consider: participants, interventions and comparisons, outcomes, resource and policy implications.	+ -	

Appendix 3: Additional Items for Quality Evaluation

Mark as 'no' if insufficient information given, and use explanatory note to indicate this.

Use "?" if there is uncertainty over the adequacy / effectiveness of the component and add explanatory note.

Engagement

1. Assessment of task engagement (e.g. evidence of boredom, frustration, off-task behaviour):
   - No assessment / assessment
   - If assessment: suggested poor engagement / suggested good engagement
   - If no assessment, any comments on likely levels of engagement
2. Assessment of whether teaching materials of interest to participants:
   - No assessment / assessment
   - If assessment: suggested low interest / suggested good interest
   - If no assessment, any comments on likely levels of interest
3. Any specific measures taken to try to enhance task engagement and avoid boredom:
   - Materials deliberately chosen as likely to be of interest to participants: Yes / No
   - Other steps taken to enhance engagement and avoid boredom (e.g. short sessions, breaks): Yes / No

Reinforcement of correct responses

4. Correct response was intrinsically reinforcing (i.e. the correct response leads to a reinforcing consequence that would occur in the participant's everyday natural environment if they used perspective-taking skills) : Yes / No
5. If artificial reinforcement was used, was an assessment conducted to ensure that the consequences were, in fact, reinforcing for the participants involved (e.g. assessment of reinforcer preferences)?
   - No assessment / assessment
   - If assessment: suggested low reinforcement / suggested good interest
   - If no assessment, any comments on likely levels of interest
6. Was reinforcement delivered immediately following a correct response? Yes / No

Feedback

7. Was clear and immediate feedback given about whether response was correct / incorrect? Yes / No
8. Was there a clear procedure for eliciting the correct response in the event of an error? Yes / No

Prompts

9. Were prompts used to increase the probability of correct responding? Yes / No
10. Was there a clear strategy for the withdrawal of the prompts? Yes / No

Generalisation and maintenance

11. Did teaching take place in the participant's everyday environment when opportunities naturally arose that required the use of deictic skills? Yes / No
12. Was there an assessment of whether the acquired skill was maintained over time (i.e. follow-up assessment)? Yes / No
13. If follow-up measures taken, were they taken after an interval of at least 2 months? Yes / No
14. Was there an assessment of the generalisation of the skill to a different context?
    - Yes / No
    - If yes:
        i. Change of context is minor / superficial (e.g. different teacher but everything else stays the same)
        ii. Change of context is complex (e.g. different task materials)
    - If yes, was there evidence of generalisation? Yes / No
15. Was there an assessment of whether the skill generalized to everyday life (i.e. whether participant was engaging in more perspective-taking (or other hypothesized consequences of deictic skills) after training?
    - Yes / No
    - If yes, was there evidence of change? Yes / No
16. Were multiple exemplars used to teach the skill in order to enhance generalisation (e.g. variation of teachers, variation of task materials, variation of places and contexts in which teaching took place)
    - Yes / No
    - If yes:
        i. Variation was just in one dimension
        ii. Variation was in several dimensions

Appendix 5: Inclusion and Exclusion Criteria

Inclusion criteria

For inclusion of the study, participants will need the following criteria:

- Individuals referred to Coventry and Warwickshire Mind or Kaleidoscope Birmingham for accessing psychological services and experiencing symptoms of depression.
- Age 18 years or older.

Exclusion criteria

People will be excluded from the study for the following reasons:

- Individuals who are not able to read and respond to questionnaires written in English.
- Individuals who have significant cognitive impairment or who do not have the capacity to provide informed consent.
- Individuals who are currently displaying active symptoms of major mental health difficulties.

Address: Hawthorns House Halfords Lane West Bromwich West Midlands **Email:** info@kaleidoscopeplus.org.uk **Telephone**: 0121 565 5605 **Website:** http://www.kaleidoscopeplus.org.uk/	

<u>Samaritans</u>

Samaritans provide confidential emotional support to anyone experiencing feelings of distress or despair. They can be contacted by phone 24 hours a day.

Address: 13 Bow Street Birmingham, West Midlands **Email:** jo@samaritans.org **Telephone:** 116 123 **Website:** http://www.samaritans.org/	

Participant Identification Number:...............

CONSENT FORM

Title of Project: Evaluation of implicit self-referential thoughts in people with depression and their association with cognitive fusion.

Researcher: Sarah Berger and Dr. Richard Bennett.

Email: SJB267@bham.ac.uk   Please tick each box if you agree:

1. I have understood the participant information sheet dated .... (version ...) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

2. I understand that my participation is voluntary and that I am free to withdraw at any time during the research, without giving any reason, without my medical care or legal rights being affected.

3. I understand that any information that could identify me will be kept strictly confidential and that no personal information will be included in the study report or other publication.

4. I understand that following the research I will have up until September 2017 when I can contact the researcher and withdraw my data should I wish. After this time data will be analysed, so removing data at this point would have an impact on the results of the research. Therefore, following this period, it will not be possible to withdraw my data from the study.

5. 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.

6. 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 tick the box ☐

Appendix 9: Procedure for calculating D scores

1, Only response-latency data from test blocks are included

2, Latencies above 10,000 ms from the data set are removed

3, Participants who produce more than 10% for test-block trials with latencies less than 300 ms are removed

4, Twelve standard deviations for the four trial-types are computed; four from the response latencies for test-blocks 1 and 2, four from test-blocks 3 and 4, and a further four from text-block 5 and 6

5, Twenty-four mean latencies for the four trial-types in each test-block are calculated

6, Difference scores are calculated for each of the four trial-types for each pair of test-blocks by subtracting the mean latency of the consistent block from the mean latency of the corresponding inconsistent block

7, Each difference score is divided by its corresponding standard deviation from step 4, producing 12 D scores, one score for each trial-type for each pair of test blocks

8, Four overall trial-type D scores, are calculated by averaging the scores for each trial-type across the three pairs of test blocks. The four trial-type scores for each participant are then used to calculate mean D scores across a group of participants