THE DEVELOPMENT AND VALIDATION OF THE COMPREHENSIVE ASSESSMENT OF ACCEPTANCE AND COMMITMENT THERAPY PROCESSES (CompACT)

ASHLEY WILLIAM FRANCIS, BSc, MSc.

Thesis submitted in part fulfilment of the requirements for the

degree of Doctor of Clinical Psychology

to the University of Nottingham

DECEMBER 2016

Acknowledgements

Firstly, I would like to thank Dr David Dawson and Dr Nima Golijani-Moghaddam, not only for pitching this idea to me as a feasible and intriguing research project after so many previous projects proved unviable, but also for their constant encouragement and enthusiasm; without their passion for research, combined with their ability to instil confidence, I am not sure that even the opportunity to receive a clinical doctorate was enough of a reinforcer to get me across the finish line!

I would also like to thank my wife Vicky, who has been able to empathise with the trials and tribulations of undertaking doctoral research, having gone through the process concurrently with me. I hope our new found freedom will see us explore the wider world beyond research and work.

Special thanks to my fellow trainees for being able to offer various opportunities for avoidance and procrastination at much needed times; I'm thinking of you Joe and Barnie in particular! Thank you also to my family for instilling in me the kind of hard work ethic that has got me this far.

Last, but by far not least, this thesis is dedicated to the memory of my sister-in-law Lizzy Strong who passed away in August 2015. Her spirit, humour and sheer bloody mindedness in the face of the greatest adversity one can ever know was extraordinary, and is still something I cannot quite understand.

Table of contents

SYSTEMATIC LITERATURE REVIEW 11
JOURNAL ARTICLE 52
Abstract 54
1. Introduction 55
1.1. Acceptance and Commitment Therapy 55
1.2. Problems with current measures of ACT processes 56
1.2.1. Specific versus general measures. 56
1.2.2. Limitations of the current general measure of ACT processes. 56
1.2.3. Six versus three core ACT processes 57
1.3. Aims and objectives 57
2. Method 58
2.1. Phase one: Measurement development 58
2.1.1. Determining what is to be measured. 58
2.1.2. Generation of initial item pool: Selecting measures and items. 58
2.1.4. Review of initial item pool by experts. 63
2.1.5. Results of the Delphi process. 64
2.2. Phase two: Testing the dimensionality, validity and reliability of the CompACT in a non-clinical sample 64
2.2.1. Participants. 64
2.2.2. Measures. 65
3. Results 69
3.1. Examination and elimination of poor, overlapping and redundant items 69
3.2. Removing multivariate outliers 69
3.4. Choosing which items to exclude/retain 71
3.5. Reliability of the CompACT 73
3.6. Validity of the CompACT 73
3.6.1. Convergent validity. 73
3.6.2. Discriminant validity. 74
3.6.3. Concurrent validity. 74
4. Discussion 76
4.1. Future research 82
5. Conclusion 82

References 83
EXTENDED PAPER 93
1. Extended introduction 94
1.1. The development of ACT 94
1.2. RFT and ACT's theory of psychopathology 96
1.3. The clinical focus of ACT 97
1.4. ACT's hexaflex model and the six core processes 98
1.4.1. Acceptance - experiential avoidance. 99
1.4.2. Defusion - fusion. 101
1.4.3. Self as context - self as content. 103
1.4.4. Present moment awareness - present moment absence 105
1.4.5. Values. 110
1.4.6. Committed action - inaction and impulsivity. 112
1.6. ACT's evidence base 114
1.7. Rationale for the current study 117
1.7.1. Limitations of the current general measure of ACT processes 117
1.7.2. Six versus three core ACT processes. 119
2. Extended methods 120
2.1. Phase one: Measurement development 120
2.1.1. Generation of initial item pool: Selecting measures 120
2.1.2. Generation of initial item pool: Item selection procedure. 121
2.1.3. Review of initial item pool by experts. 125
2.1.4. The Delphi method: A review of a consensus building approach. . 125
2.1.5. Selecting ACT experts from the ACBS website. 126
2.1.6. Selecting experts for Delphi studies. 126
2.1.7. Attrition of participants in the Delphi phase. 127
2.1.8. Rating items and providing feedback. 128
2.1.9. Consensus definition and the Delphi process. 128
2.1.10. Results of the Delphi. 130
2.1.11. Testing the readability of items in the measure 132
2.2. Phase two: Testing the dimensionality, validity and reliability of the CompACT in a non-clinical sample 134
2.2.1. Participants. 134
2.2.2. Measures: Convergent and discriminant validity. 135

2.2.3. Measures: Concurrent validity. 136
3. Extended results 138
3.1. Issues with multicollinearity 138
3.2. Identification of the CompACT's factor structure 139
3.2.1. Adequacy of sample size for factor analysis. 139
3.2.2. Extraction method. 140
3.2.3. Selection of rotation method. 140
3.2.4. Criteria determining the number of factors to extract. 141
3.2.5. Derivation of three factor structure for the CompACT. 142
3.3. Choosing which items to exclude/retain 142
4. Extended discussion 143
4.1. Strengths of the Delphi methodology 143
4.2. Factors determining the composition of the CompACT's items 144
4.3. Effects of negatively valenced items of a measure's factor structure 145
4.4. Concerns regarding the amalgamation of particular ACT processes 146
4.5. Importance of capturing values in ACT process measures 147
4.6. Further evidence to attest to the CompACT's discriminant validity from distress measures 147
4.7. Distinguishing ACT processes from symptoms of psychological distress as operationalised in self-report measures 148
5. Critical reflections 152
5.1. ACT's underpinning philosophy of science 152
5.4.2. Development as a researcher. 158
POSTER 225

List of tables and figures

Table 1 Psychometric properties of 11 ACT process measures used to extract items for 106-item pool 59

Table 2 A table illustrating the demographic details of a non-clinical sample of participants (N=377) completing the CompACT and other measures 65

Table 3 Measures used to assess different aspects of the CompACT's construct validity 67

Table 4 The CompACT's 30 items and the factors onto which they load 72

Table 5 Correlations between the CompACT, the CompACT's subscales, the AAQ-II and other measures (N= 352) 75

Table 6 Proposed key methodological criteria to report in publications of Delphi-Consensus Studies 129

Table 7 The reading ease, grade level and text statistics calculations for the CompACT 133

Figure 1 ACT's hexaflex model (adapted from Hayes et al., 1999). 99

Figure 2 Hayes et al.'s (2011) clustering of ACT processes into three dyadic groups that make up psychological flexibility. 120

List of appendices

Appendix A Ethical approval from University of Lincoln School of Psychology Research Ethics Committee (SOPREC).....190

Appendix B Initial pool of 106 items extracted from existing ACT process measures with factor loadings.....192

Appendix C Draft email to participants targeted for recruitment to phase one Delphi study.....198

Appendix D A screen print showing the Likert Scale upon which Delphi participants had to rate items from different ACT process measures.....200

Appendix E Example summary of both, the items removed, and participant's feedback in relation to items in Round one of the Delphi study.....202

Appendix F Participant information sheet for the validation of the CompACT in phase two.....207

Appendix G The CompACT's 37 items meeting threshold criteria for being retained (i.e. a rating of 'good' or 'excellent' by ≥70% of participants) following consensus being reached by participants in Round 3 of the Delphi study.....211

Appendix H A list of all measures used to assess the different aspects of the CompACT's validity.....214

Appendix I An initial Principle Axis Factoring analysis on the CompACT's initial 37 items (N= 352) with an initial four-factor solution.....218

Appendix J The seven items removed from the CompACT due to insufficient factor loadings based on Tabachnick and Fidell's (2007) (loadings <.45 on all factors; cross-loading >.2 on more than one factor).....220

Appendix K Stable factor loadings of CompACT's 30 items following a Principle Axis Factoring analysis, specifying a three factor solution and using a direct oblimin rotation method.....222

Portfolio abstract

Valid and reliable measures of Acceptance and Commitment Therapy's (ACT) purported mechanism(s) of change are paramount in establishing its incremental validity above other extant psychological therapies. However, the most widely used general measure of psychological flexibility and the six ACT processes of which it is composed of, the Acceptance and Action Questionnaire (AAQ-II), has been significantly criticised, principally in terms of being a conflated distress measure due to having items with poor face and content validity. Theoretically, there has also been debate about whether ACT's 'hexaflex' model, which conceptually underpins the AAQ-II and other ACT process measures, is comprised of fewer than six distinct processes as per current ACT theory. The overall aim of our study was therefore to: 1) develop a new general measure of ACT processes; 2) test the measure's validity and reliability; and 3) establish whether the measure's factor structure accords with Hayes et al.'s (2011) clustering of ACT's six processes into three dyadic processes. The measure was named the Comprehensive assessment of ACT processes ('CompACT').

Using a Delphi-consensus methodology, ACT experts (*N*= 10) rated the face and content validity of 106 items. Items were taken from existing ACT process measures and two mindfulness measures, as well as generated by ourselves and the ACT experts within our study. Items which failed to reach pre-defined ratings thresholds were removed and remaining items were re-rated until a pre-defined consensus threshold was reached. After three rounds, an initial 37-item version of the CompACT was produced. The CompACT was then disseminated to a non-clinical sample of participants (*N* = 377) mainly via social media. Participants also completed other measures to assess different aspects of the CompACT's validity.

A three-factor structure was found to be the most stable for a 30-itemed version of the CompACT. An examination of the distribution and contents of item loadings also confirmed that the CompACT's three factors could be defined in accordance with Hayes et al's. (2011) three dyadic ACT processes. The CompACT performed in theoretically predicted ways against measures of other related and distinct constructs, and was mostly comparable to the AAQ-II in terms of its magnitude of associations with these other measures, with the exception of a distress measure. The CompACT's associations with these measures also supported ACT's central hypothesis: higher levels of psychological inflexibility are associated with higher levels of distress and lower levels of health and wellbeing.

Overall, the CompACT shows initial promise as a comprehensive measure of ACT processes, and more generally as a measure of psychological flexibility. The psychometrically viability of Hayes et al.'s (2011) three dyadic ACT processes as a conceptual model for the CompACT further strengthens previous suggestions that ACT's clinical model may be composed of fewer than six distinct processes. The CompACT was adjudged by ACT experts to possess items with better face and content validity than the AAQ-II. However, in order to be considered as a viable alternative measure to the AAQ-II, further testing of the CompACT's psychometric properties is needed.

Statement of contribution

I declare that this research is the product of my own original work conducted since my commencement of the Trent Doctorate in Clinical Psychology training programme in 2013. Conceptually, the idea for the project was developed by Dr David Dawson and Dr Nima Golijani-Moghaddam, who also provided regular guidance and supervision with all aspects of the research, including feedback on draft versions of my thesis and journal paper, as well as help with data analysis and refining my interpretations and understanding of the implications of the data. As the Principal Investigator for this research, I have taken responsibility for: obtaining ethical approval; writing the systematic literature review; advertising the research; recruiting participants; collecting data; scoring questionnaires; entering data; undertaking the bulk of the data analysis with some additional support from Dr Nima Golijani-Moghaddam; and writing the thesis.

SYSTEMATIC LITERATURE REVIEW

Investigating the reliability and validity of the short form chronic pain acceptance questionnaire (CPAQ-8): A systematic review

Abstract

This review examined evidence pertaining to the reliability and validity of the short-form chronic pain acceptance questionnaire (CPAQ-8), and compared its psychometric properties to the full CPAQ. The methodological quality of eight studies involving treatment seeking and non-treatment seeking participants was reviewed using the COSMIN Checklist (COnsensus-based Standards for the selection of health status Measurement INstruments). Evidence suggests the CPAQ-8 retains the validity, reliability and factor structure of the full CPAQ despite being a briefer measure. Futures areas of research are suggested to improve the psychometric properties of the CPAQ-8 and enhance its utility in both clinical and research settings.

Keywords

Chronic illness, pain, pain acceptance, reliability, validation

Introduction

Modern pain management services have undergone a transition from a biological to a biopsychosocial understanding of chronic pain (CP) (Dahl et al., 2005). This has led to the development of alternative and psychological treatment approaches to the extant medical and pharmacological management of CP. One psychological approach that has developed empirical support as an effective treatment for CP is Acceptance and Commitment Therapy (ACT; Hayes et al., 1999).

ACT falls under the rubric of cognitive behavioural approaches (CBT), but diverges significantly from traditional CBT's attempts to change the negative content of thoughts in order to mediate distressing emotional and behavioural consequences (McCracken and Gutiérrez-Martínez, 2011). ACT argues that such approaches may actually increase the (experiential) avoidance of such internal experiences which then perpetuates psychological distress (Hayes et al., 1999). Instead, ACT focuses on changing how a person responds to aversive internal experiences (thoughts, feelings, physical sensations etc.), so that these experiences less impinge on the person's attempts to live a rich and meaningful life. When applied to CP, ACT suggests that it is the individual's conditioned response to avoid, minimise and/or remove pain and pain related thoughts, emotions and sensations which exacerbates the initial distress related to pain through loss of contact with valued aspects of life (McCracken and Vowles, 2014). Thus, ACT interventions centre on increasing an individual's ability to *accept* unpleasant thoughts, feelings, memories and sensations associated with pain, in order to *commit* to and pursue more valued ways of living.

ACT is supported as an effective treatment for CP (Dahl et al., 2004; McCracken et al., 2013; Thorsell et al., 2011; Wicksell et al., 2013). Evidence shows that ACT's core process of acceptance is a key predictor of emotional, social, and physical functioning in CP sufferers, with greater acceptance correlating with better functioning independent of a person's level of pain (McCracken and Eccleston, 2005). Valid and reliable measure of ACT's processes, in particular acceptance of CP, are therefore fundamental to contributing to this evidence and assessing treatment efficacy.

The Chronic Pain and Acceptance Questionnaire (CPAQ; McCracken, Vowles, et al., 2004) is a 20-item measure of acceptance in CP. Confirmatory factor analyses (CFA) have shown that two ACT processes are central in understanding the psychological aetiology of CP: (1) *activity engagement* (AE) is the process of engaging in everyday activities of value to the individual in the presence of pain; and (2) *pain willingness* (PW) is the process of disengaging from the struggle to limit contact with pain. ACT theory predicts that higher levels of CP acceptance (i.e. greater AE and PW) result in lower levels of emotional distress, and higher levels of quality of life and functioning. Whilst a previous review (see Renemen et al., 2010) identified the CPAQ to have the best internal consistency (α=.78-.91), construct validity and reliability compared to other CP measures, evidence suggests homogeneity and redundancy amongst some of the CPAQ's items (Fish et al., 2010). In combination with recommendations for briefer assessment tools in healthcare settings and internet based research in order to improve response rates (Lindell and Whitney, 2001), the CPAQ-8 (Fish et al., 2010) was developed as a short-form version of the CPAQ. This review aims to systematically evaluate studies pertaining to the CPAQ-8, and establish whether it retains the psychometric robustness of its parent measure, whilst benefiting from being a more concise tool.

Methods

Search terms

The following databases were searched simultaneously through Ovid to find studies relating to the CPAQ-8: Embase 1974 to 2016 May 24; Ovid MEDLINE (R) In Process and Non-Indexed Citations 1946 to May 24 2016; and PsychINFO 1806 to May Week 4 2016**.** Each native database was also searched to ensure that the simultaneous search reliably elicited all possible papers. Articles referencing the CPAQ-8 development article were also identified through the 'cited by' function in Google Scholar, and through the 'times cited' function in Web of Science. *'Chronic Pain* AND *Acceptance'* were entered into Ovid as a free-text search terms; *'Chronic Pain Acceptance Questionnaire'* was entered into Web of Science as a free-text search term; and ***'****Chronic Pain Acceptance Questionnaire (CPAQ-8)*' was entered into Google Scholar as a free-text search term. The search was carried out from the 1st May 2016 to 24th May, 2016.

Study selection

An initial search retrieved 1020 results. Duplicates were removed and searches were limited to articles written in English, peer reviewed, and published between 2010 and 2016[1]. 213 articles were excluded as they were either: reviewing an alternative measure of CP acceptance; were unpublished dissertations; and/or were protocols for prospective studies. The titles, abstracts and references of the remaining 230 articles were screened and 17 articles were initially identified as pertaining to the aims of this review. Hand searches of the reference sections failed to identify any additional relevant studies. Full texts of the remaining

[1] This range of dates was selected to capture articles published since the CPAQ-8 developmental study by Fish et al. (2010).

articles were accessed and screened for reference to the CPAQ-8. Articles which may not have explicitly discussed the psychometric properties of the CPAQ-8 but which allowed this information to be extracted upon analysis were included. This resulted in eight articles to be included within this review.

Data extraction

The COSMIN Checklist (COnsensus-based Standards for the selection of health status Measurement INstruments; Mokkink et al., 2010) was used to: (1) examine the quality of the available evidence for the psychometric properties of the CPAQ-8; and (2) evaluate the methodological quality of studies. The COSMIN provides raters with 114 items that are broadly grouped into reliability, validity and responsiveness. The quality of evidence for each measurement property is rated along a four-point scale (‘excellent’; ‘good’; ‘fair’; or ‘poor’) based on COSMIN criteria. A quality score per measurement property is obtained by taking the lowest rating of any item (i.e. 'worst score counts' method). The COSMIN was chosen based upon its reported reliability and validity (for further details see McCracken et al., 2010) and also because of its modular approach to the assessment of the quality of a study, which allowed studies not reporting all possible reliability and validity data for the CPAQ-8 to be rated.

Results

INSERT TABLE 1 HERE

Characteristics of the studies

See Table 1 for details of the characteristics of studies. Eight studies are included within this review (total N= 3275; mean N= 467.9; *SD*= 344; range= 120-914); four studies were cross-sectional in design; three studies were longitudinal in design; and one study was a descriptive content analysis of CP acceptance measures. Participants were predominantly female (71.8%) with a mean age of 48.6 years (*SD*= 6.27). Only four studies presented data relating to the primary cause of pain. Aetiological categories were both broad (e.g. degenerative) and specific (e.g. fibromyalgia), and thus the ability to compare these studies and report on the most common primary cause of pain is difficult. Studies consisted of treatment and non-treatment seeking participants recruited from a number of settings including: pain management programs; pain outpatient clinics; CP support groups; online CP forums; a sports centre; and CP rehab centres. One study had no participants as it sought to perform a content analysis of different measures of CP acceptance to establish whether the content of these measures' items reflected acceptance, and to what extent this was the case.

INSERT TABLE 2 HERE

Internal consistency

See Table 2 for a summary of the CPAQ-8's psychometric properties reported by studies. The CPAQ-8's internal consistency ranged between α= .78-.85 for the total scale, and between α= .79-.89 and α= .69-.77 for the AE and PW subscales respectively. Based on Terwee et al.'s (2007) suggestion of a positive rating when α= .70-.95, the CPAQ-8 and its subscales demonstrated acceptable consistency. Moreover, the CPAQ-8's internal consistency is comparable to the full CPAQ (α= .78-.91; Reneman et al., 2010) and other CP acceptance

questionnaires, including the Illness Cognitions Questionnaire (α= .84-.91; Lauwerier et al., 2010), the Pain Solutions Questionnaire (α=.78; De Vlieger et al., 2006), and the Acceptance of Illness Scale (α=.78; Rankin and Holttum, 2003).

Similarities between the CPAQ-8 and the CPAQ-8's internal consistency further extend to the characteristics of their PW subscales, which demonstrate lower levels of internal consistency compared to the AE subscale. Fish et al. (2013) argued that the negative wording of items in the PW subscale (in comparison to the AE subscale's positively keyed items) may be affecting its internal consistency. Incorporating items of opposing valence within a single measure can undermine the resolution of a stable factor structure, specifically because this engenders confusion and inaccurate responding in comparison to items posed in the same direction (van Sonderen et al., 2013). Negatively valenced items can also load onto a separate factor unrelated to the construct(s) of interest due to sharing a common variance (DiStefano and Motl, 2006). Calls for clarity in the wording of the PW subscale have also extended to the name of the subscale itself, with some arguing that reference to 'controlling and avoiding pain' may be more accurately described as 'pain unwillingness' if adhering to an ACT conceptual framework (Fish et al., 2010).

Reliability

Evidence of the CPAQ-8's test-retest reliability is limited to one study. Studies where the CPAQ-8 was administered on at least two different occasions are compromised in their ability to report test-retest reliability data due to the implementation of an intervention between data points. Fish et al. (2013) used intraclass correlation coefficients (ICCs) to find a test-retest reliability of .81, 95% CI (.73–.86) for the total CPAQ-8, .68, 95% CI (.56–.77) for the PW subscale, and .86, 95% CI (.80–.90) for the AE subscale. The total CPAQ-8 and AE

subscale's reliability coefficients are above the recommended minimum value for acceptable test-retest reliability (.70; Nunnally and Bernstein, 1994). Whilst the PW subscale (as well as other items on the AE subscale) are below this value, the CPAQ-8's overall test-retest reliability scores are comparable to those reported for the CPAQ (.79-.83; Ning et al., 2008; Ojala et al., 2013; Rodero et al., 2010).

Content validity

The CPAQ-8 was developed from its parent measure based on factor loadings (i.e. removal of items with factor loadings <.50) rather than theoretical criteria. The tool may therefore lack adequate breadth of coverage of the construct acceptance (Baranoff et al., 2014a). In the CPAQ-8 development study, Fish et al. (2010) asked three expert raters to evaluate the superiority of either item 1. or 2. for inclusion in the CPAQ-8. This was based on CFA evidence of item redundancy between these items. Unfortunately, it is unclear from Fish et al.'s reporting as to: (1) what credentials these raters had which constituted their 'expert' status; (2) what criteria they utilised to judge the suitability of items; and (3) whether raters examined other items within the CPAQ for inclusion/exclusion in the CPAQ-8. There also appears to be no involvement of CP sufferers in the development study of the CPAQ-8, which could have helped to ensure that items were comprehensible to the measure's target population. Conversely, during the construction of an Indonesian version of the CPAQ-8 for older adult CP sufferers, Lubis et al. (2013) consulted their target population by giving elderly 'non-psychology' participants the CPAQ-8 to check for readability. However, the specificity of this sample precludes any generalisability of findings to other populations, and to English language versions of the CPAQ-8. The authors also make reference to the involvement of a translator within this study, but limited information is provided which negates an analysis of the quality of this process.

Lauwerier et al. (2015) conducted a content analysis of the CP acceptance measures to ascertain how and to what extent these measures assess acceptance in CP. Within their sample of measures, all iterations of the CPAQ were examined including: the CPAQ-8; CPAQ; CPAQ adolescent version (CPAQ-A; Wallace et al., 2011) and the original CPAQ-34 (Geiser, 1992). The authors applied a coding framework underpinned by two key theoretical accounts of CP acceptance, ACT and self-regulatory theory (Carver and Scheier, 1982). This framework helped to analyse and categorise each measures assessment of CP acceptance in three key domains: (1) *disengagement from pain control* (letting go of attempts to control pain); (2) *willingness to experience pain* (without attempting to reduce, avoid, or otherwise change pain); and (3) *engagement in activities other than pain control*. The authors also defined additional categories unrelated to CP acceptance including: *controlling pain*; *pain costs; pain benefits; unclear* (i.e. no clarity regarding what the content of an item relates to)*;* and *no fit with any categories.* There was moderate general agreement amongst raters regarding their assessments of the CPAQ-8's item content (62.5%). Just under half of the CPAQ-8's items (46.25%) were considered to assess acceptance. However, this was only in the domain of *engagement in activities other than pain control*. Raters also considered that two-fifths (40%) of the CPAQ-8's items examined additional categories other than pain acceptance, including *controlling pain* (37.5%) and *pain costs* (2.5%). The CPAQ-8's remaining items were rated as either *unclear* as to what categories they fit into (3.75%), or they were rated as *not fitting into any of the categories* (10%). The CPAQ and CPAQ-34 (Geiser, 1992) were also considered to have items which mostly reflected either *engagement in activities other than pain control* (42%; 30.83%) or *controlling pain* (35%; 25%). Lauwerier et al. (2014) noted however that the underrepresentation of other important areas of acceptance within the CPAQ-8's items is shared by other CP acceptance measures. Additionally, multidimensional scaling within this study found *pain willingness* and *engagement in activities other than pain control* as the two key dimensions capturing the

content of the 112 items across 13 CP acceptance measures. This supports ACT's conceptualisation of CP acceptance as being constructed of these two factors, and whilst the CPAQ-8 purports to reflect these aspects of acceptance within its PW and AE subscales, Lauwerier et al. (2014) suggest that only the activity engagement aspect of acceptance is evident within the CPAQ-8's items. Again however, the majority of CP acceptance measures were also considered not to assess these features of acceptance simultaneously.

All other iterations of the CPAQ-8 were rated as having at least some items which reflected the other domains of acceptance. In particular, the original CPAQ-34 had the highest (but still a small) number of items representing *disengagement from pain control* (14.58%) and *pain willingness* (14.58%) compared to other versions of the CPAQ. Lauwerier et al. (2014) suggested that the process of refining and reducing the number of items within the CPAQ from its original 34 items to 8 items in the CPAQ-8 may have negatively impacted on the breadth of acceptance content within the measure. The authors also criticise all iterations of the CPAQ as well as other CP acceptance measures for having negatively valenced items, as such items arguably measure CP 'non-acceptance' rather than acceptance. The authors suggest some problems which may emerge from this include: obfuscating what the measure is actually assessing for clinicians and researchers (e.g. attempts to avoid or control pain, but not willingness to experience pain may be erroneously framed within a fear-avoidance paradigm); and creating 'false' or inflated negative correlations between outcomes such as pain willingness and avoidance of pain and acceptance and disability.

Structural validity

McCracken et al. (2004) first identified that CP is comprised of activity engagement and pain willingness. A number of CFAs have subsequently confirmed the goodness-of-fit of a two factor structure underpinning the CPAQ (Fish et al., 2010; van Tulder et al., 1997; Vowles et al., 2008; Wicksell et al., 2009). Overall, the CPAQ-8 appears to have retained the two factor structure of the CPAQ. However, findings are more mixed in relation to which measure possesses the greater fit to this structure.

Baranoff et al. (2014b) found the CPAQ-8 provided an overall better fit than the CPAQ, which failed to evidence an acceptable well-fitting two factor model based on absolute close-fit indices. Rovner et al. (2014) also found the CPAQ-8 had a better fit in all evaluated goodness of fit indices for a two factor structure compared to the CPAQ. This appears to relate to the exclusion of two items (4,16) from the CPAQ PW subscale which had factor loadings of ≤0.40 (thus not achieving acceptable limits). The CPAQ's item 16 has previously been critiqued for its poor wording (Wicksell et al., 2013) and its exclusion from the CPAQ-8 increased the factor loadings onto the PW subscale (0.51-0.82). This suggests that item content/wording may have a significant bearing on factor structure. Conversely, Fish et al. (2013) found that a one factor model had a similarly good fit to a two factor structure for the CPAQ-8. Other researchers have further questioned the validity of a two factor structure, suggesting that the wording effects implicit within these measures may be producing a falsified two factor structure (Ning et al., 2008). However, Fish et al. provide a number of reasons as to why the CPAQ-8 is better accounted for by a two versus a single factor structure beyond goodness of fit statistical evidence, including: failure to replicate a one factor in the validation sample; moderate intercorrelation (r= .38) between AE and PW subscales suggesting a degree of discriminant validity; greater explanation of the variance of criterion variables following the

addition of the PW subscale after the AE subscale (Reneman et al., 2010); and differential relationships of the subscales with criterion variables.

Hypotheses testing

All derivations of the CPAQ are underpinned by ACT principles, which seek to enhance acceptance, reduce avoidance and increased engagement with valued activities. Therefore, ACT related measures such as the CPAQ-8 should: capture changes in acceptance; correlate with other empirically established pain-mediating process variables; and correlate with other outcome measures targeting key pain-related attributes in the expected direction.

Consistent with findings from studies involving the CPAQ (Wicksell et al., 2013), the CPAQ-8 correlated with process measures that have been found to mediate the impact of CP on depression and disability (van Sonderen et al., 2013). These include; catastrophising (*r*= -.53) and pain intensity (*r*= -.53) (Baranoff et al., 2014b); psychological inflexibility in pain (*r*= -.78) and pain self-efficacy (*r*= .72) (Fish et al., 2013); and fear of movement/re-injury (kinesiophobia) (*r*= -.57) (Reneman et al., 2010)[2]. Comparison of the CPAQ and CPAQ-8's correlations with pain-related outcome variables also suggests theoretically consistent relationships of a similar magnitude with: the number of medical visits for pain (*r*= -.33) (Fish et al., 2010); pain intensity (CPAQ *r*= -.23; CPAQ-8 *r*= -.23); depression (*r*= -.54; -.48); anxiety (*r*= -.36; *r*= -.31); stress (*r*= -.44; *r*= -.38) and disability (*r*= -.40; *r*= -.42) (Baranoff et al., 2014b).

[2] Some of these studies examined the CPAQ-8 in relation to multiple process measures, but for the sake of clarity and to avoid repetition, they are presented as covering separate process measures from other studies within this review.

Rovner et al. (2015) conducted a latent class analysis (LCA) of the CPAQ and CPAQ-8. Both measures were found to have a 4-cluster structure based on a sample of 914 chronic pain patients. Membership to each cluster was determined by variations in pain acceptance (i.e. PW and AE) resulting in four subgroups: (1) 'low' acceptance group (low PW and AE); (2) 'medium' acceptance group (medium PW and AE); (3) 'high' acceptance group (high PW and AE); and (4) 'mixed' acceptance group (low PW and high AE). This suggests that the CPAQ and CPAQ-8 are comparable in their ability to distinguish between patients in terms of degrees of pain acceptance. Emotional functioning (Hospital Anxiety and Depression Scales (HADS); Zigmond and Snaith, 1983), pain-related fear (Tampa Scale for Kinesiophobia (TSK); Miller et al., 1991) and physical functioning (Multidimensional Pain Inventory (MPI); Bergström et al., 1998) were also found to be significantly associated with pain acceptance in theoretically predicted ways; patients categorised by the CPAQ-8 as being in the high acceptance group were found to have lower levels of emotional distress and higher levels of physical functioning than patients in the mixed, medium and low acceptance groups.

In some cases the CPAQ-8 was found to explain more of the variance in depression and disability when controlling for catastrophising and kinesiophobia (Baranoff et al., 2014a), and pain intensity (Baranoff et al., 2014b). This suggests that the CPAQ-8 is tapping into a construct (acceptance) that is distinct, and is thus adding incremental validity to established pain-mediating process measures within the literature. However, such findings have not been completely reflected in other studies. Baranoff et al. (2014b) found equivalent variance between the CPAQ-8 and kinesiophobia to disability outcomes, and catastrophising contributed a larger amount of unique variance to depression than the CPAQ-8. Despite this, the authors re-assert the argument for CPAQ-8's incremental validity by showing its

statistically significant and greater contribution to disability when catastrophising is accounted for.

Discrepancies were found in the strength of correlation between the CPAQ-8's subscales and outcome measures. The majority of studies indicated that the AE subscale has a larger effect size for all outcome variables than the PW subscale. An exception to this was PW exhibited a stronger correlation with cognitive fusion (*r*= -.43) and fear of pain (*r*= -.45) than AE (*r*= -.22; -.32) (Fish et al., 2013). These findings replicate those of the original CPAQ, which was criticised due to the PW subscale being less robust than the AE scale (Nicholas and Asghari, 2006; Wicksell et al., 2013). However, McCracken et al. (2004) suggest that the AE and PW subscales are oriented towards measuring different facets of pain consistent with the underlying ACT model, and therefore should correlate differently with different outcomes. Additionally, the stronger correlation between the AE subscale and outcomes may also relate to the previously discussed effects of negative wording on the PW subscale. Although this would again imply that the PW subscale is flawed, evidence indicates that the CPAQ-8's AE and PW subscales together fully mediate the impact of pain severity on outcomes, but when separated only provide partial mediation (Fish et al., 2013).

Cross-cultural validity

The CPAQ has been validated and translated when applied to other cultures and languages, such as Swedish (Wicksell et al., 2013), Spanish (Rodero et al., 2010), German (Nilges et al., 2007) and Chinese (Ning et al., 2008). The CPAQ-8 has also been found to retain the psychometric properties and two factor structure of its parent measure when tested in different countries, including Ireland (Fish et al., 2010, 2013), Australia (Baranoff et al., 2014a, 2014b) and Sweden (Rovner et al., 2014). However, the relative infancy of the CPAQ-8

means it has only been translated into two other languages; Lubis et al. (2013) back-translated the CPAQ-8 into Indonesian but did not subject the measure to a rigorous validation process; Rovner et al. (2014) extracted the eight items from the previously validated Swedish CPAQ and found evidence of generalisability of the dimensionality, internal consistency (α≥.80) and reliability findings of the CPAQ-8 to a Swedish population and language.

Criterion validity

Strong correlations (*rs*= .92-.93) between the CPAQ and CPAQ-8 are exhibited across all studies, with the CPAQ-8 also able to account for between 80 to 86.5% of the total variance of the CPAQ, and 85% of the variance for the CPAQ total scale score (Baranoff et al., 2014b; Fish et al., 2010; Reneman et al., 2010). This suggests that the CPAQ-8 is measuring the same underlying construct as its parent measure, despite possible threats to the measures construct validity through item exclusion. Both measures also converge in terms of a stronger relationship between their AE subscale and criterion variables than the PW subscale, with the exception of one study which showed approximately equivalent correlations between the PW and AE subscales and criterion variables for both the CPAQ (*rs*= -.15-.49) and the CPAQ-8 (*rs*= -.19-.49) (Baranoff et al., 2014b).

Fish et al. (2013) used different comparators in order to establish the construct validity of the CPAQ-8. A weak to moderate correlation (*r*= .48) was found between the CPAQ-8 and a validated general measure of acceptance (Acceptance and Action Questionnaire (AAQ-II); Bond et al., 2011). This finding may reflect a divergence in terms of the dimensionality of these measures, with the AAQ-II found to be underpinned by a single factor structure of

psychological flexibility, and the CPAQ/CPAQ-8 underpinned by a two factor structure. Fish et al. also found the CPAQ-8 to have a strong negative correlation (*r*= -.78) with the Psychological Inflexibility in Pain Scale (PIPS; Wicksell, Renöfält, et al., 2008). This relationship is consistent with the ACT literature, which suggests commonality between the measures in terms of their focus on acceptance processes, but also difference in terms of the PIPS' focus on cognitive fusion processes which have been linked to avoidance of pain and pain-related internal experiences (Wetherell et al., 2011).

One study found the CPAQ-8 to be the only process measure to predict reliable deterioration in a composite of depression and/or disability (as measured by the Reliable Change Index; (Jacobson and Truax, 1991). Whilst no other studies investigated the predictive criterion validity of the CPAQ-8 in relation to key pain outcome measures, this data suggests a possibility utility of the CPAQ-8 as a post-treatment monitoring tool of changes in depression and disability connected to CP.

Responsiveness

Three studies within this review investigated longitudinal changes (6-9 weeks) in pain-related outcome variables using the CPAQ-8 following psychological intervention for CP. These studies found that the CPAQ-8 had a moderate effect size for sensitivity to changes in treatment targets, including disability (*rs*= -.53-.55) and depression (*rs*= -.48-.51). Based on the CPAQ's reported sensitivity to changes in disability (*r*= -.37) and depression (*r*= -.31) (McCracken and Gutiérrez-Martínez, 2011), the CPAQ-8 appears to be marginally more sensitive. However, Rovner et al. (2014) found the CPAQ had a slightly larger effect size (*r*= .33) for sensitivity to outcome changes following pain rehabilitation than the CPAQ-8 (*r*= .27).

Together, these findings indicate a likely comparability between the CPAQ-8 and CPAQ in terms of responsiveness.

Discussion

The CPAQ-8 was developed due to evidence of homogeneity and redundancy amongst some of the CPAQ items, and also to reduce participant burden in completing the measure. Findings of the current review overall suggest that the CPAQ-8 retains the reliability, dimensionality and validity of the CPAQ. However, these findings are mixed and require further discussion in light of methodological flaws of some of the studies, as well as the paucity of research pertaining to the CPAQ-8.

The CPAQ-8 demonstrated comparable internal consistency to the full 20-item CPAQ across all studies. Given that a greater number of test items can inflate alpha values (Cortina, 1993), this finding is perhaps surprising but may relate to the CPAQ-8's exclusion of items from the CPAQ with low item-to-factor loadings (<.50) and low item-total correlations (<.50). Whilst high alpha values may also suggest item homogeneity and redundancy (Boyle, 1991), the satisfactory high item factor loadings of the CPAQ-8 (.51-.84) indicate that the measure is internally coherent and valid.

Whilst the CPAQ-8 and its AE subscale demonstrated test-retest reliability comparable to that the CPAQ, evidence relating to this was limited to one study and thus conclusions about this aspect of the measure must be tempered. Moreover, the CPAQ's PW subscale fell below recommended levels, possibly due to a number of reasons which will require further research

(e.g. impact of negative wording on subscale; unaccounted for changes between test administrations etc.). It is important to note that test-retest reliability relies on the premise of stability in the underlying construct being assessed between successive administrations of a measure. The CPAQ-8/CPAQ are underpinned by a functional contextualist perspective and suggest that 'acceptance' has a state-like quality that is shaped and altered by context; as such, it is not expected to be stable, but changeable and highly context dependent. Although the CPAQ-8/CPAQ must demonstrate sound repeatability to be of potential pragmatic use (e.g. monitoring changes in acceptance following CP treatment), it may be more apt to consider the utility of functional contextual measures in terms of responsiveness to change, rather than constancy of scores across administrations (Golijani-Moghaddam et al., 2013). Given the CPAQ-8's small to moderate effect size for detecting post-treatment longitudinal changes in depression and disability, the tool demonstrates potential utility as a post-treatment progress measure, but more research is need to strengthen this conclusion.

Some concerns are apparent over the CPAQ-8's ability to broadly measure acceptance due to the selection of its items according to factor loadings versus theoretical criteria. The lack of consultation with experts in the field of acceptance and/or CP during the development of the CPAQ-8 also serves to further compound these concerns. Lack of involvement of the target population is also limited. Given that the comprehensibility of the CPAQ-8's negatively worded PW subscale is suggested to be a key mediator of the scales reliability, validity and structure, a lack of information in this regard from target populations is problematic. There is also a suggestion that the removal of items from the original 34-item version to the current 8-item CPAQ has impacted on the breadth of acceptance content within the measure, such that only the *engagement in activities other than pain control* aspect of this construct is represented. Moreover, a third of the CPAQ-8's items were considered to measure *controlling pain* which

does not directly relate to acceptance, but may indirectly do so (i.e. by controlling pain one is less likely to 'disengage from pain control' or 'engage in activities other than pain control'). This criticism was also levelled at other iterations of the CPAQ and all other CP acceptance measures assessed; again, this may reflect issues relating to the preponderance of negatively valenced items within these measures which assess CP acceptance indirectly. Despite this issue, all studies found the CPAQ-8 retained the two factor structure of the CPAQ and thus coheres with the underpinning ACT theory that CP acceptance is comprised of activity engagement and pain willingness. However, mixed findings were evident in relation to which CPAQ had best fit overall, with no overall consensus evident.

The CPAQ-8 was found to cohere with the CP acceptance literature by correlating with pain-related outcomes, process variables and other ACT-related measures in theoretically consistent ways. Evidence in these studies show that as pain acceptance increases, ratings of depression and disability decrease. Acceptance was also able to account for more of the variance in depression and disability than other factors (even after accounting for catastrophising and kinesiophobia). This further indicates that acceptance is an important and unique mediating variable of CP on emotional and physical functioning. Correlations found between the CPAQ other ACT-related process measures further validates the measure's consistency with the underlying ACT model. Conceptual and linguistic commonalities amongst ACT process measures exist, meaning that correlations between the measures may be artificially inflated once participants become socialised to them (i.e. learning effects). However, this issue appears to be balanced out by the weight of other evidence within this review which confirms the adherence of the CPAQ-8 to the underlying ACT model.

Structural equation modelling showed that the CPAQ-8's AE and PW subscales only partially mediate the impact of pain intensity on depression and disability when analysed separately. This strengthens the theoretical argument that acceptance in CP requires both AE and PW, despite the latter's inferior predictive validity of key outcomes (McCracken, 2010). McCracken stresses that whilst an analysis of the differential contributions of the CPAQ-8's subscales is replete within the literature, individual examination and comparison of the constituents of acceptance may lead to an important loss of understanding acceptance as a whole process.

The ability of the CPAQ-8 to explain more of the variance in depression and disability post CP treatment than other process measures, as well as being the only predictor of reliable deterioration in pain-related outcomes, also supports the measures predictive criterion validity. The CPAQ-8 was also found to be comparable to the CPAQ in its ability to differentiate patients according to their levels of pain acceptance, which meaningfully correlated with emotional and physical functioning outcomes, as per ACT theory. Importantly however, the CPAQ-8's predictive criterion validity was established in most studies through correlations with other self-report measures. This raises a potential issue of common method variance (see Lindell and Whitney, 2001). Therefore, studies may have benefited from examining the CPAQ-8 in relation to behavioural/observational/or other non-self-report measures.

Cross-cultural applicability of the CPAQ-8 has been demonstrated by confirmation of its reliability, validity and responsiveness in Irish, Australian and Swedish populations, and also when translated into Swedish. Comparatively, whilst the CPAQ has been tested in more countries and more languages, the relative infancy of the CPAQ-8 means this is an area of future development to increase the measure's uptake globally.

INSERT TABLE 3 HERE

Limitations

Issues relating to the methodological quality of each study can be seen in Table 3. Only eight studies investigated the CPAQ-8's measurement properties directly, or indirectly. Thus, there is a limited amount of data attesting to the psychometric properties of the CPAQ-8. Of particular concern is: the (mostly) poor ratings for evidence of content validity and responsiveness; fair to poor ratings for cross cultural validity; and fair ratings for hypotheses testing. In some cases, these ratings appear to accurately reflect important methodological weaknesses including: failure to describe a time interval between administrations of the CPAQ-8 (Reneman et al., 2010) and what happened between the time period (Lubis et al., 2013; Reneman et al., 2010); failure to describe how missing data was handled (Fish et al., 2013); and a failure to stipulate explicit a priori hypotheses about the nature and strength of expected relationships between acceptance and pain-related outcome variables (Baranoff et al., 2014b; Fish et al., 2010, 2013; Reneman et al., 2010). However, some of these ratings also appear to reflect the overly stringent quality criteria promoted by the COSMIN, which has been critiqued elsewhere (Reneman et al., 2010). COSMIN quality criteria was developed through a Delphi consensus style study, in response to dissatisfaction with the variation in standards of research (Terwee et al., 2012). Whilst these criteria may reflect 'best practice', it has been argued that such high standards do not reflect current accepted practices in research and when applied, would exclude much methodologically sound evidence (Angst, 2011). Although the COSMIN standards within this review helped to illustrate how far from

agreed ideals the current evidence is (with scope to inform future research), it is possible that the use of other quality criteria protocols may have evaluated studies more favourably. In turn, this would have allowed stronger conclusions to be drawn. Moreover, the categories of reliability and validity defined by the COSMIN were too broad at times to be of use and thus, more specific categories may have helped to elucidate all facets of reliability and validity more clearly.

Another important caveat when reviewing the COSMIN's quality ratings of studies is in the relation to the aims of studies. Lubis et al. (2013) aimed to pilot CBT therapy for Indonesian older adults with a range of problems including CP rather than to investigate the psychometric properties of the CPAQ-8. Thus when judged against the COSMIN's quality criteria for assessment of psychometric properties, the study is deemed methodologically poor. However, when measured against its actual aims, the study can be rated as more methodologically sound despite limitations remaining (e.g. small sample size).

Implications

Aside from limited data pertaining to the CPAQ-8, as well as some of the methodological flaws in the studies reported, evidence indicates that the shorter form measure benefits from increased brevity without significant loss of any of the precision of the full CPAQ. Potential implications of this include: better completion rates amongst participants/service-users in research and clinical settings where testing burden can be high; and more time to administer other measures of other important variables granting researchers the opportunity to test out alternative hypotheses.

The finding that changes in acceptance mediates the impact of pain intensity on mood and disability beyond catastrophising and kinesiophobia, reinforces the clinical importance of treatments which target acceptance either directly (e.g. ACT) or indirectly (e.g. CBT) (Vlaeyen and Linton, 2012). This also provides theoretical support for pain acceptance as a unique process variable in CP.

Future research

Future research may benefit from stipulating a priori hypotheses about the nature and strength of relationships expected between the CPAQ-8 and other measures. This would allow comparison of the observed and expected findings, and allow conclusions to be drawn either confirming hypotheses or, highlighting issues with the CPAQ-8's reliability (i.e. if the observed effects are too small or large based on theoretically informed hypotheses). Given that the comprehensibility of the CPAQ-8's PW subscale has been found to be a key mediator of the scale's psychometric properties, future research may benefit from investigating this through consultation with CP sufferers. Consultation with experts in the field of acceptance and/or CP in a Delphi consensus type study may also help to confirm whether the content of the CPAQ-8 is broad enough to fully measure acceptance. As Fish et al. (2013) note, future research may also examine the effects of wording by adapting the CPAQ-8 such that all items are phrased unidirectionally (i.e. all positively or negatively worded), and then compare this to the original CPAQ-8 in terms of their respective psychometric properties. Finally, evidence for the CPAQ-8's construct validity may further be enhanced by examining whether the CPAQ can discriminate between known groups (i.e. those high and low in CP-related depression and disability), and whether the CPAQ-8 converges with behavioural/observational/or other non-self-report measures of pain-related constructs.

Summary and conclusions

The CPAQ-8 demonstrates comparable psychometric properties to its parent measure despite being a more parsimonious tool. Clinical implications relating to the utility of the CPAQ-8 as a post-treatment monitoring tool follow from this research, but must be tempered based on the limited data and methodological issues. Recommendations are made for areas of improvement in future research which may help to increase the CPAQ-8's profile amongst research and clinical communities.

Declaration of conflict of interests

None declared.

References

Angst F (2011) The new COSMIN guidelines confront traditional concepts of responsiveness. *BMC Medical Research Methodology* 11(1): 152-158.

Baranoff J, Hanrahan SJ, Kapur D, et al. (2014a) Six month post-treatment deterioration in acceptance (CPAQ-8) and cognitions following multidisciplinary pain treatment. *Journal of Behavioral Medicine* 37(3): 469–79.

Baranoff J, Hanrahan SJ, Kapur D, et al. (2014b) Validation of the Chronic Pain Acceptance Questionnaire-8 in an Australian pain clinic sample. *International Journal of Behavioral Medicine* 21(1): 177–85.

Bergström G, Jensen IB, Linton SJ and Nygren AL (1999) A psychometric evaluation of the Swedish version of the Multidimensional Pain Inventory (MPI-S): A gender differentiated evaluation. *European Journal of Pain* 3:261-273.

Bond FW, Hayes SC, Baer RA, 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.

Boyle GJ (1991) Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and individual differences* 12(3): 291–294.

Carver CS and Scheier MF (1982) Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin* 92(1): 111-135.

Cortina JM (1993) What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology* 78(1): 98-104.

Dahl J, Wilson KG and Nilsson A (2004) Acceptance and commitment therapy and the treatment of persons at risk for long-term disability resulting from stress and pain symptoms: A preliminary randomized trial. *Behavior Therapy* 35(4): 785–801.

Dahl J, Luciano C and Wilson KG (2005) *Acceptance and commitment therapy for chronic pain*. Reno, NV: New Harbinger Publications.

De Vlieger P, Bussche E Van Den, Eccleston C, et al. (2006) Finding a solution to the problem of pain: Conceptual formulation and the development of the Pain Solutions Questionnaire (PaSol). *Pain* 123(3): 285–293.

DiStefano, C., and Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, *13*(3), 440-464.

Fish RA, Hogan MJ, Morrison TG, et al. (2013) Willing and able: A closer look at pain willingness and activity engagement on the Chronic Pain Acceptance Questionnaire (CPAQ-8). *The Journal of Pain* 14(3): 233–45.

Fish RA, McGuire B, Hogan M, et al. (2010) Validation of the chronic pain acceptance questionnaire (CPAQ) in an Internet sample and development and preliminary validation of the CPAQ-8. *Pain* 149(3): 435–43.

Geiser DS (1992). *A comparison of acceptance focused and control-focused psychological treatments in a chronic pain treatment center*. PhD Thesis, University of Nevada, Reno, USA.

Golijani-Moghaddam N, Hart A and Dawson DL (2013) The implicit relational assessment procedure: Emerging reliability and validity data. *Journal of Contextual Behavioral Science* 2(3): 105–119.

Hayes SC, Strosahl KD and Wilson KG (1999) *Acceptance and commitment therapy: An experiential approach to behavior change*. New York, NY: The Guildford Press.

Jacobson NS and Truax P (1991) Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology* 59(1): 12-19.

Kratz AL, Davis MC and Zautra AJ (2007) Pain acceptance moderates the relation between pain and negative affect in female osteoarthritis and fibromyalgia patients. *Annals of Behavioral Medicine* 33(3): 291–301.

Lauwerier E, Caes L, Van Damme S, et al. (2015) Acceptance: What's in a name? A content analysis of acceptance instruments in individuals with chronic pain. *The Journal of Pain* 16(4): 306-317.

Lauwerier E, Crombez G, Van Damme S, et al. (2010) The construct validity of the illness cognition questionnaire: The robustness of the three-factor structure across patients with chronic pain and chronic fatigue. *International Journal of Behavioral Medicine* 17(2): 90–96.

Lindell MK and Whitney DJ (2001) Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology* 86(1): 114-121.

Lubis DU, Jaya ES, Arjadi R, et al. (2013) Preliminary study on the effectiveness of short group cognitive behavioral therapy (GCBT) on Indonesian older adults. *PLoS ONE* 8(2).

McCracken LM (2010) Toward understanding acceptance and psychological flexibility in chronic pain. *Pain* 149(3): 420–421.

McCracken LM and Eccleston C (2005) A prospective study of acceptance of pain and patient functioning with chronic pain. *Pain* 118(1): 164–169.

McCracken LM and Gutiérrez-Martínez O (2011) Processes of change in psychological flexibility in an interdisciplinary group-based treatment for chronic pain based on Acceptance and Commitment Therapy. *Behaviour Research and Therapy* 49(4): 267–274.

McCracken LM and Vowles KE (2014) Acceptance and commitment therapy and mindfulness for chronic pain: Model, process, and progress. *American Psychologist* 69(2): 178-187.

McCracken LM, Carson JW, Eccleston C, et al. (2004) Acceptance and change in the context of chronic pain. *Pain* 109(1): 4–7.

McCracken LM, Vowles KE and Eccleston C (2004) Acceptance of chronic pain: Component analysis and a revised assessment method. *Pain* 107(1): 159–166.

McCracken LM, Vowles KE and Zhao-O'Brien J (2010) Further development of an instrument to assess psychological flexibility in people with chronic pain. *Journal of Behavioral Medicine* 33(5): 346–354.

McCracken LM, Sato A and Taylor GJ (2013) A trial of a brief group-based form of acceptance and commitment therapy (ACT) for chronic pain in general practice: Pilot outcome and process results. *The Journal of Pain* 14(11): 1398–1406.

Miller RP, Kori SH and Todd DD (1991). The Tampa Scale. *Unpublished report*. Tampa, FL.

Mokkink LB, Terwee CB, Gibbons E, et al. (2010) Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. *BMC Medical Research Methodology* 10(1): 82

Nicholas MK and Asghari A (2006) Investigating acceptance in adjustment to chronic pain: Is acceptance broader than we thought? *Pain* 124(3): 269–279.

Nilges P, Köster B and Schmidt CO (2007) Pain acceptance-concept and validation of a German version of the chronic pain acceptance questionnaire. *Schmerz* 21(1): 57–58.

Ning MC, Ming TWC, Mae JYC, et al. (2008) Validation of the chronic pain acceptance questionnaire (CPAQ) in Cantonese-speaking Chinese patients. *The Journal of Pain*, 9(9): 823–832.

Nunnally JC and Bernstein IH (1994) *Psychometric theory*. New York, NY: McGraw-Hill.

Ojala T, Piirainen A, Sipilä K, et al. (2013) Reliability and validity study of the Finnish version of the Chronic Pain Acceptance Questionnaire (CPAQ). *Disability and Rehabilitation* 35(4): 306–314.

Rankin H and Holttum SE (2003) The relationship between acceptance and cognitive representations of pain in participants of a pain management programme. *Psychology, Health and Medicine* 8(3): 329–334.

Reneman MF, Dijkstra A, Geertzen JHB, et al. (2010) Psychometric properties of Chronic Pain Acceptance Questionnaires: a systematic review. *European Journal of Pain* 14(5): 457–65.

Rodero B, García-Campayo J, Casanueva B, et al. (2010) Research validation of the Spanish version of the Chronic Pain Acceptance Questionnaire (CPAQ) for the assessment of acceptance in fibromyalgia. *Health and Quality of Life Outcomes* 8: 37.

Rovner GS, Arestedt K, Gerdle B, et al. (2014) Psychometric properties of the 8-item Chronic Pain Acceptance Questionnaire (CPAQ-8) in a Swedish chronic pain cohort. *Journal of Rehabilitation Medicine* 46(1): 73–80.

Rovner G, Vowles KE, Gerdle B and Gillanders D (2015) Latent class analysis of the short and long forms of the Chronic Pain Acceptance Questionnaire: Further examination of patient subgroups. *Journal of Pain* 16(11):1095-105.

Terwee CB, Bot SDM, De Boer MR, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 60(1): 34–42.

Terwee CB, Mokkink LB, Knol DL, et al. (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research* 21(4): 651–657.

Thorsell J, Finnes A, Dahl J, et al. (2011) A comparative study of 2 manual-based self-help interventions, acceptance and commitment therapy and applied relaxation, for persons with chronic pain. *The Clinical Journal of Pain* 27(8): 716–723.

van Sonderen E, Sanderman R and Coyne JC (2013) Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloS one* 8(9).

van Tulder MW, Koes BW and Bouter LM (1997) Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions. *Spine* 22(18): 2128–2156.

Vlaeyen JWS and Linton SJ (2012) Fear-avoidance model of chronic musculoskeletal pain: 12 years on. *Pain* 153(6): 1144–1147.

Vowles KE, McCracken LM, McLeod C, et al. (2008) The Chronic Pain Acceptance Questionnaire: Confirmatory factor analysis and identification of patient subgroups. *Pain* 140(2): 284–291.

Wallace DP, Harbeck-Weber C, Whiteside SPH, et al. (2011) Adolescent acceptance of pain: Confirmatory factor analysis and further validation of the chronic pain acceptance questionnaire, adolescent version. *The Journal of Pain* 12(5): 591–599.

Wetherell JL, Afari N, Rutledge T, et al. (2011) A randomized, controlled trial of acceptance and commitment therapy and cognitive-behavioral therapy for chronic pain. *Pain* 152: 2098–2107.

Wicksell RK, Renöfält J, Olsson GL, et al. (2008) Avoidance and cognitive fusion–central components in pain related disability? Development and preliminary validation of the Psychological Inflexibility in Pain Scale (PIPS). *European Journal of Pain* 12(4): 491–500.

Wicksell RK, Olsson GL and Melin L (2009) The Chronic Pain Acceptance Questionnaire (CPAQ)-further validation including a confirmatory factor analysis and a comparison with the Tampa Scale of Kinesiophobia. *European Journal of Pain* 13(7): 760–768.

Wicksell RK, Kemani M, Jensen K, et al. (2013) Acceptance and commitment therapy for fibromyalgia: A randomized controlled trial. *European Journal of Pain* 17(4): 599–611.

Zigmond AS and Snaith RP. (1983) The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*. 67(6):361-70.

Table 1.

Characteristics of the studies in the systematic literature review.

Author/year/location	Sample size (n)	Gender M/F (%)	Age mean yrs (SD)	Primary cause of pain (%)	Study design	Setting/method of participant selection
Baranoff et al. (2014a) Australia.	120	45/55	43.1 (9.7)	Degenerative (51.4) Traumatic/surgical (31.6) Genetic (7.5) Psychological (0.9)	Longitudinal	Participants had completed end of program assessment and 6-month follow-up assessment following a 3 week MDT CBT-based pain management program.
Baranoff et al. (2014b) Australia.	334	42.6/57.4	46.2 (11.9)	Work-related (65) Motor-vehicle (12) Unknown (14) Other (10)	Cross-sectional	Participants were recruited from an outpatient pain clinic in a public hospital.
Fish et al. (2013) Ireland.	550 (478 IN sample; 72 PP sample). 105 TRT sample.	IN: 19.5/80.5 PP: 26.8/73.2 TRT: 14.4/85.6	IN: 42.8 (12.3) PP: 52.6 (14.1) TRT: 47.3 (10.36)	Arthritis (IN 7.2; PP 7.0 TRT 7.9) Fibromyalgia (31.3; 11.1; 42.9) Migraine (3.5; 6.9; 1.0) Nerve damage (14.1; 27.8; 16.2) Disc problems (7.4; 18.1; 1.0) Sciatica (1.6; 0; 0) Postsurgical pain (2.0; 0; 2.9) Traumatic injury (8.6; 5.6; 8.6) Neuropathy (2.0; 6.9; 2.9)	Longitudinal	CP participants recruited online and from a CP support organisation. 105 completed CPAQ-8 again within 6 weeks providing test-retest data.

Author/year/location	Sample size (n)	Gender M/F (%)	Age mean yrs (SD)	Primary cause of pain (%)	Study design	Setting/method of participant selection
Fish et al. (2010) Ireland.	428 (319 IN sample; 109 PP sample).	IN: 20.1/79.6 PP: 16.8/83.2	IN: 44.6 (12.6) PP: 54.1 (12.1)	Arthritis (IN 35.1; PP 29.8) Fibromyalgia (42.6; 22.8) Migraine (27.6; 15.8) Nerve damage (32; 47.4) Disc problems (27.6; 34.2) Sciatica (12.5; 9.6) Traumatic injury (14.1; 20.2)	Cross-sectional	Participants (n= 319) recruited through online CP discussion groups and forums which hosted questionnaire. Also recruited from the ICPA and local pain and physiotherapy practices (n= 109).
Lauwerier et al. (2014) Belgium.	-	-	-	-	Descriptive	The authors conducted a content analysis of different chronic pain acceptance measures, including the CPAQ-8.
Lubis et al. (2013) Indonesia.	38	29/71	63.4 (7.25)	No data provided on participant's primary cause of pain.	Cross-sectional	Older adult participants recruited from sports centre. Inclusion required meeting cut-off points following completion of battery of questionnaires and consent to treatment. Participants divided 4 treatment groups including CP.
Rovner et al. (2014) Sweden.	891	33.9/66.1	47.5 (14)	Primary cause of pain not given. Instead, area of pain location and % given (e.g. head and face (5.7)).	Longitudinal	891 participants recruited from a rehabilitation centre. 91 of these participants underwent ACT-based rehab, supplying responsiveness data for CPAQ-8.

Author/year/location	Sample size (n)	Gender M/F (%)	Age mean yrs (SD)	Primary cause of pain (%)	Study design	Setting/method of participant selection
Rovner et al. (2015) Sweden	914	34.1/65.9	47.5 (14.7)	Widespread pain (26.8) Lower back (15.4) Neck (14.9) Hips/lower limbs (13.5) Shoulders/upper limbs (12.8) Neck and face (5.7) Abdomen/sexual organs (4.1) Upper back (2.5)	Cross-sectional	914 patients recruited via a pain rehabilitation centre.

Abbreviations. CP (Chronic Pain), MDT (Multi-disciplinary team), CBT (Cognitive Behavioural Therapy), ICPA (Irish Chronic Pain Association), CPAQ-8 (Chronic Pain Acceptance Questionnaire-Version 8), IN (Internet sample), TRT (Test-rest sample) PP (Paper and Pencil sample).

Table 2.

Reviewed studies with empirical data pertaining to the reliability and/or validity of the chronic pain acceptance questionnaire (CPAQ-8).

Author/date	Internal consistency	Reliability	Content validity (including face validity)	Structural validity (Construct validity)	Hypotheses testing (Construct validity)	Cross-cultural validity (Construct validity)	Criterion validity	Responsiveness
Baranoff et al. (2014a)	CPAQ-8 total *α*= .85	?	?	?	CPAQ-8 explained more of the variance in depression and disability (depression *r*= -.44; disability *r*= -.22) beyond catastrophising (*r*= .35; *r*= .04) and kinesiophobia (*r* = -.26; *r*= .31).	Evidence of underlying two factor structure for CPAQ-8 in Australian sample.	CPAQ-8 only significant predictor of reliable change in depression and/or disability (*r*= -.16).	Changes in CPAQ-8 correlated with concurrent changes in depression (*r*= -.53*)* and disability (*r*= -.48).
Baranoff et al. (2014b)	CPAQ-8 total *α* =.80. AE subscale *α* = .82. PW subscale *α* = .77.	?	?	CFA found support for a 2-factor model (pain willingness, activity engagement) for CPAQ-8.	CPAQ-8 correlated with pain intensity (*r*= -.23), catastrophising (*r*= -.53) and kinesiophobia (*r*= -.53) and mood outcomes (*rs*= -.31 -.48). CPAQ-8 subscales accounted for more of variance in depression and disability (depression *rs*= -.21-.32; disability *rs*= -.15 - .26) beyond pain (*rs*= .19; .33).	Evidence of underlying two factor structure for CPAQ-8 in Australian sample.	CPAQ-8 and CPAQ correlated (*r*= .93).	

Author/date	Internal consistency	Reliability	Content validity (including face validity)	Construct validity (structural validity)	Construct validity (hypotheses testing)	Construct validity (cross-cultural validity)	Criterion validity	Responsiveness
Fish et al. (2013)	CPAQ-8 total *α*= .79 -.82 across two samples. AE subscale *α*= .85 -.86. PW subscale *α*= .69 -.74.	ICC's show test-retest reliability of CPAQ-8 (*r*= .81) 4-6 weeks apart.	?	2-factor structure for CPAQ-8 confirmed in validation sample.	CPAQ-8 correlated with emotional distress (*r*= -.38-.63), pain severity (*r*= -.27) and pain process variables (*rs*= -.46 -.79). PW stronger correlation with cognitive fusion (r= -.43) and fear of pain (*r*= -.45) than AE (*r*= -.22; -.32).	?	CPAQ-8 correlated with AAQ-II (*r*= .48) and PIPS (*r*= -.79).	?
Fish et al. (2010)	Total α= .78-.85. AE subscale *α*= .79 -.89. PW subscale *α*= .77.	?	Items removed from CPAQ based on <.50 item total correlations and factor loadings.	CFA suggests adequate to good fit of a two factor structure underpinning the CPAQ-8.	CPAQ-8 correlated with number of medical pain visits (*r*= -.33), pain interference (*r*= -.44), pain severity (*r*= -.38), depression (r= -.59) and anxiety (*r*= -.69).	?	CPAQ-8 and CPAQ were strongly correlated (*r*=.81).	?
Lubis et al. (2013)	?	?	Examined adapted Indonesian CPAQ-8 for face validity using two clinical psychologist specialists in Gerontology. Gave adapted CPAQ-8 to elderly people to scrutinise for readability.	?	?	CPAQ-8 adapted through back-translation for Indonesian older adults. No CFA to check underlying structure of adapted scale.	?	CPAQ-8 scores sensitive to pre to post changes in depression (r= -.55) following GCBT treatment for chronic pain.

Author/date	Internal consistency	Reliability	Content validity (including face validity)	Construct validity (structural validity)	Construct validity (hypotheses testing)	Construct validity (cross-cultural validity)	Criterion validity	Responsiveness
Rovner et al. (2014)	*α*= .80	?	?	CFA showed CPAQ-8 had better fit in all fit indices than the CPAQ, whilst still retaining a 2-factor structure.	CPAQ-8 correlated with: anxiety (*r* = -0.40), depression (*r* = -.47), kinesiophobia (*r* = -.57) and quality of life subscales (*rs* = .35 -.48).	Evidence of underlying two factor structure for Swedish CPAQ-8.	CPAQ-8 and CPAQ correlated (*r*= .92).	Small effect sizes found for CPAQ-8 (r= .27) and CPAQ (*r*= .30) in relation to sensitivity to longitudinal changes in outcome measures.
Rovner et al. (2015)	?	?	?	?	Patient self-ratings on the CPAQ-8 correlated in theoretically predicted ways with emotional and physical functioning outcome measures. Patients in high pain acceptance group had lower levels of anxiety, depression, pain related fear, and higher levels of general activity than medium and low pain acceptance groups.	?	?	?

Note. All data reported is statistically significant (p<0.05). Effect sizes (0.1 - 0.3 [weak], 0.4 - 0.6 [moderate], 0.7 - 1.0 [strong]) are reported as Pearson's product moment coefficients (*r*) for ease of comparability across studies, but may have been reported as standardised Beta (*β*) coefficients or Cohen's *d* in the original studies and have been subject to conversion.

Abbreviations: CPAQ-8 (Chronic Pain acceptance Questionnaire-Version 8), CPAQ (Chronic Pain Acceptance Questionnaire-original 20 item version), CBT (Cognitive behavioural Therapy), MDT (Multi-disciplinary team), AAQ-II (Acceptance and Action Questionnaire - version II), PIPS (Psychological Inflexibility Scale), AE (Activity Engagement subscale), PW (Pain Willingness subscale), CFA (Confirmatory Factor Analysis), α (Cronbach's alpha coefficient), ICC (Intraclass Correlation Coefficients), ? (no information available).

Table 3.

Methodological quality assessment and investigated measurement properties per study according to the COSMIN checklist (COnsensus-based Standards for the selection of health status Measurement INstruments).

Author/date	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypothesis testing	Cross-cultural validity	Criterion validity	Responsiveness
Baranoff et al. (2014a)	Good	?	?	?	?	Fair	?	Fair	Fair
Baranoff et al. (2014b)	Good	?	?	?	Good	Fair	?	Good	?
Fish et al. (2013)	Excellent	Good	?	?	Excellent	Fair	?	Fair	?
Fish et al. (2010)	Excellent	?	?	Poor	Excellent	Fair	?	Excellent	?
Lauwerier et al. (2014)	?	?	?	Excellent	?	?	?	?	?
Lubis et al. (2013)	?	?	?	Poor	?	?	Poor	?	Poor
Rovner et al. (2014)	Good	?	?	?	Good	Fair	Fair	Good	Poor
Rovner et al. (2015)	?	?	?	?	?	Fair	?	?	?

Abbreviations. ? (data was not reported or not assessed).

JOURNAL ARTICLE

The Development and Validation of the Comprehensive Assessment of Acceptance and Commitment Therapy Processes (CompACT)[3]

Ashley Francis[a*]
Email: lwxaf7@nottingham.ac.uk

[a]University of Nottingham
Division of Psychiatry and Applied Psychology
School of Medicine
Jubilee Campus
Yang Fujia Building
Wollaton Road
Nottingham, UK
NG1 8BB

Tel: (+44)1158 466646

Dr. David Dawson[b]
Email: Ddawson@lincoln.ac.uk

Dr. Nima Golijani-Moghaddam[b]
Email: NMoghaddam@lincoln.ac.uk

[b]University of Lincoln
Health, Life and Social Sciences
1st Floor, Bridge House
Brayford Pool
Lincoln, UK
LN6 7TS
Tel: (+44)1522 886029
Fax: (+44)1522 837390

*_Author note._ The study was carried out and funded as part of a doctoral course of study in clinical psychology, undertaken by the first author. Correspondence concerning this paper should be addressed to the first author: Ashley Francis, DClinPsy student, University of Nottingham, Division of Psychiatry and Applied Psychology, School of Medicine, Jubilee Campus, Yang Fujia Building, Wollaton Road, Nottingham, UK, NG1 8BB, lwxaf7@nottingham.ac.uk, (+44)1158 466646.

[3] For Submission to the Journal of Contextual Behavioural Science (JCBS). Author's guidelines (>4 printed pages) are available from: https://www.elsevier.com/journals/journal-of-contextual-behavioral-science/2212-1447/guide-for-authors.

Abstract

Valid and reliable measures of Acceptance and Commitment Therapy (ACT) processes are needed in order to establish ACT's theorised mechanisms of change over and above other extant therapies. However, existing ACT process measures are limited due to their extensive overlap and conflation with distress outcome measures that are theoretically distinct from ACT processes, or because they only examine single ACT processes and therefore do not help substantiate ACT's overall model. Conceptually, there are also questions as to whether there are fewer than six distinct processes underpinning ACT and ACT process measures. The aim was therefore to develop a new general measure of ACT processes and test the measure's factor structure, validity and reliability. In phase one, ACT experts (*N* = 10) rated the face and content validity of 106 items using a Delphi consensus methodology, and produced an initial 37-itemed Comprehensive measure of ACT processes (CompACT). In phase two, a non-clinical sample of participants (*N* = 377) completed the CompACT and measures of other related and unrelated constructs. An exploratory factor analysis suggested a three factor structure for a 30-itemed version of the CompACT in accordance with Hayes, Villatte, Levin and Hildebrandt's (2011) clustering of ACT's six processes into three dyadic processes. The CompACT also demonstrated good internal consistency, and correlated with other measures in accordance with ACT theory: higher levels of psychological inflexibility were associated with higher levels of distress and lower levels of health and wellbeing. The CompACT therefore shows initial promise as a general measure of ACT processes.

Key words: Acceptance and Commitment Therapy, psychological flexibility, assessment, measures, AAQ, hexaflex

1. Introduction

1.1. Acceptance and Commitment Therapy

Acceptance and Commitment Therapy (ACT; Hayes, Strosahl & Wilson, 1999) is a transdiagnostic approach, which conceptualises all forms of psychological distress as emerging from cognitive entanglement, experiential avoidance and values-inconsistent behaviour (see extended paper 1.1 - 1.2.). The objective in ACT is not to target distress directly, but indirectly through increasing *psychological flexibility* (the ability to contact the present moment, and to change or persist in behaviour when this serves valued ends; Hayes, Luoma, Bond, Masuda & Lillis, 2006). This indirect approach to ameliorating psychological distress by addressing the function, rather than the form of aversive stimuli (i.e. negative thoughts) is a feature of 'third wave' cognitive behavioural approaches, and represents a significant departure from historical 'first' and 'second wave' treatments (see extended paper 1.3.).

The model guiding ACT's clinical application is the 'hexaflex'. Psychological flexibility is placed centrally within the hexaflex to denote its role as *the* mechanism of change (Ciarrochi, Bilich & Godsel, 2010), and is comprised of six distinct, but interrelated processes: acceptance; defusion; self as context; present moment awareness; values; and committed action (Hayes et al., 2006; see extended paper 1.4. - 1.4.6). Psychological flexibility has been found to mediate important therapeutic outcomes, with higher levels of psychological flexibility correlating with lower levels of psychological distress and improved quality of life (Hayes, Levin, Plumb-Vilardaga, Villatte, & Pistorello, 2013). Whilst the overall evidence for ACT as an efficacious therapy remains far from definitive (see Öst, 2008; 2014 - see extended paper 1.6.), it has been argued that research should also focus on identifying the theoretical mechanisms by which treatments work (Kazdin, 2007). This may be particularly important in ACT research, given ACT's focus on changing processes rather than outcomes relating to psychological distress and behavioural functioning (Levin & Hayes, 2009). Crucially, such research is dependent on valid and reliable measurement of ACT processes.

1.2. Problems with current measures of ACT processes

1.2.1. Specific versus general measures. There has been a significant proliferation of measures that examine ACT processes in specific clinical domains including: chronic pain (e.g. Vowles, McCracken, McLeod, & Eccleston, 2008); weight-related issues (e.g. Lillis & Hayes, 2008); and diabetes management (e.g. Gregg Callaghan, Hayes & Glenn-Lawson, 2007). However, such measures are somewhat incongruent with ACT's trans-diagnostic goal (Lang et al., 2012) of targeting common processes underpinning different forms of psychological distress. There has also been a focus on developing measures of single ACT processes, including: acceptance/experiential avoidance (e.g. Gámez, Chmielewski, Kotov & Ruggero, 2014); fusion/defusion (e.g. Gillanders et al., 2014); and values (e.g. Wilson, Sandoz, Kitchens & Roberts, 2011). Although such measures might be useful to examine the differential impact of these single processes on behaviour, they do not help support the overall ACT model more broadly.

1.2.2. Limitations of the current general measure of ACT processes. The Acceptance and Action Questionnaire (AAQ-II; Bond et al., 2011) is the most widely used general measure of ACT processes (Ruiz, Herrera, Luciano, Cangas & Beltran, 2013). Despite its ubiquity, the AAQ-II has received significant criticism within the literature, mainly due to conflating ACT processes with distress outcomes such as 'worry' and 'meta-worry' (see Gámez, Chmielewski, Kotov, Ruggero & Watson, 2011; Gámez et al., 2014; Wolgast, 2014 - see extended paper 1.7.1). Theoretically these constructs are purported to be distinct from ACT processes, and the AAQ-II's difficulty in discriminating these constructs appears to derive from possessing poorly worded items which overlap with the contents of distress. Similarly, the AAQ-II's predecessor, the Acceptance and Action Questionnaire (AAQ; Hayes et al., 2006) was also criticised for having poorly worded items (see Chawla & Ostafin, 2007), and the continuing presence of this issue is both surprising and problematic, as it suggests that the AAQ-II is actually a distress outcome measure, rather than an ACT process measure. I also argue that the AAQ-II is limited in a number of other ways, including: 1) having a preponderance of items which focus on acceptance and defusion, but very few, or no items examining other important

ACT processes in the hexaflex; and 2) possessing negatively valenced items only, thus making the AAQ-II a measure of psychological in*flexibility* rather than a measure of psychological *flexibility* as it is purported to be (see 4. for a discussion on this).

1.2.3. Six versus three core ACT processes. There have been several attempts to offer a more parsimonious account of psychological flexibility comprised of fewer than six distinct processes (see extended paper 1.7.2.). These accounts suggest that each ACT process is conceptually more deeply interlinked with another, and thus processes may be combined into fewer processes (see Hayes et al., 1999; Hayes, 2004; Hayes, Villatte, Levin & Hildebrandt, 2011). More recently Hayes et al. (2011) propose that the hexaflex is actually underpinned by three 'dyadic' processes: 1) 'openness to experience and detachment from literality' (acceptance; defusion); 2) 'self-awareness and perspective taking' (present moment awareness; self as context); and 3) 'motivation and activation' (values; committed action). Blackledge and Drake (2013) suggest that translating complex lab-based RFT processes into simpler 'midlevel' clinical ACT processes has been far from seamless, and that the derivation of six distinct processes within the ACT model may be an artefact of this translation issue. They further suggest that for the ACT model to be fully consistent with its underpinning RFT account, new technical terms that subsume two or more of ACT's 'midlevel' processes may have to be developed.

1.3. Aims and objectives

The aim was to develop a general measure of ACT processes, whilst holding in mind the limitations of current ACT process measures such as the AAQ-II. I also aimed to test the factor structure, validity and reliability of this measure, and establish whether Hayes et al.'s (2011) theorised three dyadic ACT processes are psychometrically viable as a conceptual framework to underpin this measure. I named the measure the Comprehensive assessment of ACT processes ('CompACT').

2. Method

The study was granted ethical approval (see extended paper Appendix A) and encompassed two phases: phase one sought to develop the CompACT through psychometric review and the use of a Delphi methodology; phase two sought to test the CompACT's psychometric properties.

2.1. Phase one: Measurement development

2.1.1. Determining what is to be measured. Psychological theory (in this case ACT theory) provides clarity and direction in determining the breadth, content, and boundaries of the construct(s) of interest (Clark & Watson, 1995). Without such considerations there is a risk that test items may inadvertently stray into the content of other related domains. As many of the items were taken from existing measures, the breadth, content, and boundaries of each item were pre-defined by other researchers. I therefore provided our own definition of each dyadic ACT process.

2.1.2. Generation of initial item pool: Selecting measures and items. 106 mixed valence items formed the initial item pool, from which a smaller number of items could be selected to form the CompACT. Items were taken from 11 measures of both existing ACT process measures, and measures of other theoretically related constructs (the Philadelphia Mindfulness Scale [PMS; Cardaciotto, Herbert, Forman, Moitra, & Farrow, 2008] and the Mindfulness Attention and Awareness Scale [MAAS; Brown & Ryan, 2003]) (see extended paper 2.1.1). Measures were selected based on their established psychometric properties (see Table 1). Three items with the highest factor loadings from each measure were selected for inclusion in the initial item pool (see extended paper 2.1.2). Some items were also selected based on their apparent face and content validity, even if they had lower factor loadings than other items, as we had concerns about the face and content validity of some of the items despite their high factor loadings. I also generated 37 items representing different ACT processes.

Table 1

Psychometric Properties of 11 ACT Process Measures Used to Extract Items for 106-item pool

Measure	Source reference	Number of items	Description of measure and subscales	Example item	Reliability
Acceptance/experiential avoidance process measures					
Acceptance and Action Questionnaire (AAQ-II)	Bond et al. (2011)	7	Assesses psychological (in)flexibility.	I worry about not being able to control my worries and feelings.	Internal reliability: (.78-.88) Test-retest reliability: 3 months (.81) 12 months (.79)
Philadelphia Mindfulness Scale (PMS)	Cardaciotto et al. (2008)	20	Bi-dimensional measure of mindfulness, assessing it as two constructs: 1) present-moment awareness; and 2) acceptance*.	I tell myself that I shouldn't have certain thoughts.	Internal reliability: Awareness subscale (.81) Acceptance subscale (.85) Interitem correlations: Awareness subscale (.13-.50) Acceptance subscale (.17-.54)
Brief Experiential Avoidance Questionnaire (BEAQ)	Gámez et al. (2014)	15	Experiential avoidance measure of six domains: 1) behavioural avoidance; 2) distress aversion; 3) procrastination; 4)distraction/suppression ; 5) repression/denial; 6) distress endurance.	I rarely do something if there is a chance it will upset me.	Internal reliability: (.86) Mean Interitem correlation (.30)

Measure	Source reference	Number of items	Description of measure and subscales	Example item	Reliability
Defusion/fusion process measures					
Cognitive Fusion Questionnaire (CFQ)	Gillanders et al. (2014)	7	Measures cognitive fusion as a single factor.	I get upset with myself for having certain thoughts.	Internal reliability: (0.88)
Believability of Anxious Feelings and Thoughts Scale (BAFT)	Herzberg, Sheppard, Forsyth, Credé, Earleywine and Eifert (2012)	16	Measures cognitive fusion as three factors: 1) fusion with somatic concerns; 2) fusion with emotional regulation; and 3) fusion with negative evaluation.	I could lose control of myself when I feel anxious or afraid.	Internal reliability: BAFT total (.90) Somatic subscale (.81) Emotional subscale (.81) Negative evaluation subscale (.84)
Avoidance and Fusion for Youth Questionnaire (AFQ-Y)	Greco, Lambert and Baer (2008)	17	A measure of psychological inflexibility engendered by high levels of cognitive fusion and experiential avoidance.	I must get rid of my worries and fears so I can have a good life.	Internal reliability: (.90)
Drexel Defusion Scale (DDS)	Forman et al. (2012)	10	Measures cognitive fusion as a single factor.	Thoughts of self - imagine you are having a thought such as "no one likes me". To what extent would you normally be able to *defuse* from negative thoughts about yourself?	Internal reliability: (.83) Interitem correlations: (.15-.66)
Present moment awareness/present moment absence process measure					
Mindfulness Attention Awareness Scale (MAAS)	Brown and Ryan (2003)	15	Measures how inattentive and unaware an individual is of their present moment experience.	It seems I'm running on automatic pilot, without much awareness of what I am doing.	Test retest reliability: (.81) Internal reliability: (.80-.87)

Measure	Source reference	Number of items	Description of measure and subscales	Example item	Reliability
Self as context process measure					
Self as Context Scale**	Gird and Zettle (2013)	11	Assesses self as context as two factors: 1) Transcending (an enduring perspective from which experience can be observed); and 2) Centring (an ability to find stability in the face of emotional turmoil).	Despite the many changes in my life, there is a basic part of who I am that remains unchanged.	Internal reliability: (.84)
Values process measures					
Engaged Living Scale (ELS)	Trompetter et al. (2013)	16	Assesses engaged living as two constructs: 1) Valued Living; and 2) Life Fulfilment	I have values that give my life more meaning.	Internal reliability: ELS total (.90) Valued living subscale (.86) Life fulfilment subscale (.86)
Valuing Living Questionnaire	Wilson et al. (2011)	10	Assesses valued living across 10 different domains, and separates this into two factors: 1) Importance of the domain to an individual; and the 2) Consistency of a person's actions with their values in each domain.	It seemed like I was just going through the motions, rather than focussing on what was important to me.	Internal reliability: VLQ total (.65-.74) Importance subscale (.79-.83) Valued living subscale (.65-.74) Test-retest reliability: VLQ total (.57-.79) Importance subscale (.61-.82) Consistency subscale.43-.61)

Measure	Source reference	Number of items	Description of measure and subscales	Example item	Reliability
Committed action process measure					
Committed Action Questionnaire, Short Form (CAQ-8)	McCracken, Chilcot and Norton (2015)	8	Assess committed action with both positive and negatively valenced items, which create two separate factors.	I am able to follow my long terms plans including times when progress is slow.	Internal reliability: CAQ-8 total (.87) Positive subscale (.87) Negative subscale (.80)

Note. Internal reliabilities calculated using Cronbachs alpha (*α*); Test-retest reliability calculated using Pearson's (*r*) correlations coefficients.

* Items selected from the PMS were used to measure acceptance.

** Self as context items were only included for rating from round three of the Delphi phase of the study.

2.1.4. Review of initial item pool by experts. Typically, a measure's items are reviewed by experts within the field of interest to establish their face and content validity (DeVellis, 1991 - see extended paper 2.1.3.). In order to do this, a Delphi-consensus methodology was employed, which uses a series of questionnaires to gather expert opinion on a particular subject matter, with the aim of generating accord amongst those experts (Hsu & Sanford, 2007 - see extended paper 2.1.4). ACT experts from research and clinical fields were targeted for recruitment via the Association for Contextual Behavioural Science's (ACBS) website. The website contains a membership directory and member profiles, which enables individuals to document their expertise and experience in ACT (see extended paper 2.1.5.). A person was considered to be an 'expert' if they were either: 1) researchers with three or more publications relating to ACT; 2) clinicians who have an ACT professional accreditation through the British Association for Behavioural and Cognitive Psychotherapies (BABCP) or another professional body within their practising country; and/or 3) clinicians who have been using ACT as their primary model of choice in clinical practice for >5 years (see extended paper 2.1.6).

Based on recommendations (see Netemeyer, Bearden & Sharma, 2003), I decided that between 10-20 participants would be sufficient to achieve the necessary breadth of opinion in the study, whilst keeping the number of participants manageable enough to ensure adequate time for consensus building. 50 participants were targeted for recruitment to the expert panel in order to accommodate for the fact that many of these participants may not be able to participate in our study. 13 ACT experts completed the initial round of item rating, and 10 participants completed all rounds of rating; thus there was a total attrition of three participants between the first and final round (see extended paper 2.1.7.).

The panel were asked to rate each item's face and content validity along a five-point Likert scale (1= poor; 5= excellent - see extended paper Appendix E; see extended paper 2.1.8.). Items which failed to achieve either a rating of 4 ('good') or 5 ('excellent') by ≥70% of panel members were excluded from further rounds of rating. Qualitative information regarding items was also gathered from the

expert panel. This helped to establish whether the initial item pool adequately covered the breadth and scope of each ACT process. Some additional items were included for rating based on the panel's suggestions.

Best practice guidelines suggest that consensus criteria should be defined a priori (Diamond, Grant, Feldman, Pencharz, Ling, Moore, & Wales 2014; see extended paper 2.1.9.). Consensus was predefined as a Cronbach's alpha (α) agreement coefficient of ≥0.80 based on other Delphi studies (e.g. Singh, Aggarwal, Grantcharov & Darzi, 2013; Zevin, Bonrath, Aggarwal, Dedy, Ahmed & Grantcharov, 2013). When consensus was achieved, only those items rated 4 ('good') or 5 ('excellent') by ≥70% of experts in the final round were included in the CompACT.

2.1.5. Results of the Delphi process. After three rounds of rating consensus was achieved and the Delphi process completed (see extended paper 2.10.1.). The first iteration of the CompACT contained 37 items. Items were tested for comprehensibility through the Flesch Reading Ease Score test (FRES) and the Flesch Kincaid Grade Level test based on recommendations (see Streiner & Norman, 1995 - see extended paper 2.10.2). 10 community based adults unfamiliar with ACT also completed the measure and provided feedback regarding the comprehensibility of items. The final items were then sent back to the expert panel and were given final approval.

2.2. Phase two: Testing the dimensionality, validity and reliability of the CompACT in a non-clinical sample

2.2.1. Participants. I recruited a non-clinical sample of participants (N=377) primarily via social media, as well as through advertising the research at the University of Lincoln (see Table 2 for demographic details - see extended paper 2.2.1).

Table 2

A Table Illustrating the Demographic Details of a Non-Clinical Sample of Participants (N = 377) Completing the CompACT and Other Measures

	Total (N= 377) (%)
Gender	
Male	96 (25.5)
Female	279 (74)
Mean Age (years)	31.34 (*SD* = 11.12)
Ethnicity	
White	356 (94.4)
Asian	10 (2.7)
British	7 (1.9)
Mixed	3 (0.8)
Race	
Other	

Abbreviations. SD (Standard Deviation).

2.2.2. Measures. Participants completed the CompACT and ratings were made on a seven-point Likert scale from 0 ('never true') to 6 ('always true'). Participants also completed other measures (see Table 3) to assess whether the CompACT: 1) converged with measures of theoretically related constructs (AAQ-II); 2) could be discriminated from measures of theoretically divergent constructs (Short form Marlowe-Crowne Social Desirability scale [MCSD; Ballard, 1992]); and 3) correlated with measures of outcome variables in expected ways (Depression Anxiety Stress Scale [DASS-21; Henry & Crawford, 2005]; Short Form Health Survey [SF-12v2; Ware, Kosinski & Keller, 1996; Ware, Kosinski, Turner-Bowker & Gandek, 2002]) (see extended paper 2.2.2. - 2.2.3.). Based on ACT theory, it was predicted that higher levels of

psychological inflexibility should be associated with higher levels of psychological distress and lower levels of health and wellbeing.

Table 3

Measures Used to Assess Different Aspects of the CompACT's Construct Validity

Measure	Source reference	Number of items	Description of measure and subscales	Example item	Scoring	Reliability
Short form Marlowe-Crowne Social Desirability scale (MCSD)	Ballard (1992)	13	Assesses when a participant's responding is constrained by social desirability.	I sometimes feel resentful when I don't get my way.	Participants are asked to provide dichotomous 'yes' or 'no' responses to items. Higher scores indicate more socially desirable responses.	Internal reliability: (.62-.76)
Depression Anxiety Stress Scale (DASS-21)	Henry and Crawford (2005)	21	Distress measure assessing three separate constructs: 1) Depression; 2) Anxiety; and 3) Stress.	Depression item: I couldn't seem to experience any positive feeling at all.	Each item is scored from 0 (did not apply to me at all over the last week) to 3 (applied to me very much or most of the time over the past week). Higher scores indicate greater levels of distress.	Internal reliability: (.82-.97)

Short Form Health Survey (SF-12V2)	Ware et al. (1996; 2002)	12	A general measure of health and wellbeing with two factors: 1) Physical health; and 2) Mental health.	In general would you say your health is: Excellent; Very Good; Good; Fair; Poor?	Ratings on five-point Likert scale (from 'poor' to 'excellent') of their functioning over the last four weeks across eight domains:1) physical functioning; 2) role limitations due to physical problems; 3) social functioning; 4) bodily pain; 5) mental health; 6) role limitations due to emotional problems; 7) vitality; and 8) general health perceptions. Higher scores indicate better health and quality of life.	Internal reliability: Physical health scale (.92) Mental health scale (.88) Test-retest reliability: two weeks (.76-.89)

Note. Internal reliability scores are calculated using Cronbach's alpha (*α*); Test-retest reliability scores use Pearson's (*r*) correlations coefficients.

3. Results

3.1. Examination and elimination of poor, overlapping and redundant items

Corrected-item total correlations were examined to see if items fell below Nunnally and Bernstein's (1994) recommended threshold (i.e. r <.30). Items falling below this were considered to be measuring different constructs to other items, and thus would negatively impinge upon the measure's overall internal consistency. Inter-item correlations were also examined to identify redundant items (i.e. items where extensive overlap suggests they are measuring the same content and thus not adding further predictive validity; Cohen & Swerdlik, 2005). Items were considered poor and removed if either: 1) they had zero or very low correlations (r <.3) with other items; or 2) they had very high correlations (r > .8). None of the CompACT's initial 37 items fell within either of these parameters for exclusion, and thus all items were retained at this stage.

An examination of the correlation matrix revealed some evidence of multicollinearity amongst the CompACT's initial 37 items, as the determinant of the correlation matrix (.00001 [1.506E-10]) was below recommended threshold values (see Field, 2013). However, multicollinearity did not prove to be a major issue as I was able to find a stable factor structure for the CompACT (see 3.4. below; see extended paper 3.1.). Bartlett's Test of Sphericity was found to be significant (χ^2= 7640.56, df, = 666, p < .001), suggesting that the CompACT's items are indeed intercorrelated and therefore related as expected based on ACT theory.

3.2. Removing multivariate outliers

In order to ensure the CompACT was not overly influenced by the presence of outliers (i.e. responses which are extreme compared to the mean responses on at least two or more variables; Field, 2013), I examined our dataset using Mahalanobis distance values. Cases are considered to be multivariate outliers if the probability associated with their Mahalanobis distance value is p < .001. 25 cases meeting this criterion were removed and the CompACT's factor structure (and later its validity) was analysed based on the remaining 352 cases.

3.3. Identification of the CompACT's factor structure

Whilst it was hypothesized that the CompACT may best be captured by three factors, I believed that the measure's dimensionality should be determined via an inductive versus deductive process. An exploratory factor analysis (EFA) was therefore performed in the first instance (see extended paper 3.2.). Results from a Kaiser Meyer Olkin Test (Kaiser, 1970) (.92) and Bartlett's Test of Sphericity (χ^2= 7640.56, df, = 666, p<.001) suggest that our sample size (N=352) was above adequate for EFA (see extended paper 3.2.1.). Factors were extracted via Principle Axis Factoring (PAF), and an oblique rotation method (direct Oblimin) was used (see extended paper 3.2.2. - 3.2.3). Factors were extracted based on: 1) Kaiser's (1974) criteria (retain factors with an eigenvalue >1 rule); 2) a 'Scree Test' (factors are retained which precede the steep curve in the line before the first point where the line begins a flat trend; Cattell & Vogelmann, 1977); and 3) Horn's (1965) Parallel Analysis (retain factors with eigenvalues that exceed randomly generated eigenvalues at p < .05 - see extended paper 3.2.4.).

PAF on the CompACT's 37 items (N= 352) found an initial four-factor solution (see extended paper 3.2.5. and extended paper Appendix H). However, as I proceeded with a four-factor solution and began to remove poorly loading items a four-factor solution was not stable. Eight items were initially removed following the PAF with a pre-specified four-factor solution. A second PAF on the remaining 29 items was then performed, and the eigenvalues for the fourth factor (1.336) did not exceed the random eigenvalues generated by a Parallel Analysis (1.431) at this point; a three-factor solution was now implicated as the most suitable factor structure for the CompACT. I therefore re-ran the PAF analysis on the CompACT's initial 37 items and specified a three-factor solution. Three factors accounted for 50.1% of the variance and 46.2% of the cumulative variance following extraction.

3.4. Choosing which items to exclude/retain

Items with insufficient loadings onto the CompACT's three factors were removed to produce the most stable factor structure (see 3.3.). I applied Tabachnick and Fidell's (2007) criteria of removing items with: 1) loadings below .45 on all factors; and 2) cross-loading items with a difference of at least .2 between the primary factor and the cross-loading factor(s) (see extended paper 3.4). Seven items were removed leaving 30 remaining items. A PAF analysis was re-run specifying a three factor solution with the same extraction and rotation procedures. Again the KMO Test (.91) and Bartlett's Test of Sphericity (χ^2= 6323.15, df, = 435, p<.001) suggested that the sample size was adequate. This time the analysis was stable, with all items loading >.45, and no cross-loadings within .20 of a primary loading (see extended paper Appendix J). Thus, all 30 items were retained. An examination of each item's valence suggested that the majority of items were negatively valenced, with the exception of motivation-activation items, which were all positively valenced (see Table 4). A three factor solution again provided the greatest stability based on a PAF analysis, with the eigenvalue on the third factor (2.297) exceeding the random eigenvalue (1.495). A three factor solution on the CompACT's 30 items explained 54.6% of the total variance and 49.9% of the cumulative variance.

Each of the three factors were examined to see which items most strongly loaded onto, and therefore best represented, each factor (see Table 4). The content of these items was then analysed to help define the factors. Based on Henson and Robert's (2006) suggestion, factors were defined on the basis of >3 items which were theoretically derived from the same ACT process loading onto the same factor. The results suggest that every item within the CompACT loaded onto the specific factor that it was hypothesized to. The following three 'subscales' were identified and had high internal consistencies: Factor 1: 'openness to experience and detachment from literality' (11 acceptance items; 3 defusion items; Cronbach's alpha = .91); Factor 2: 'motivation and activation' (11 values/committed action items; Cronbach's alpha = .91).; and Factor 3: 'self-awareness and perspective taking' (5 contact with present moment/mindfulness items; Cronbach's alpha = .87).

Table 4

The CompACT's 30 Items and the Factors onto Which They Load

	Three Factor Solution		
	Factor 1 (α=.91)	Factor 2: (α=.91)	Factor 3: (α=.87)
1. I tell myself that I shouldn't have certain thoughts	.68		
2. I try to stay busy to keep thoughts or feelings from coming	.68		
3. I try to distract myself when I feel unpleasant emotions	.62		
4. One of my big goals is to be free from painful emotions	.74		
5. I go out of my way to avoid situations that might bring difficult thoughts, feelings, or sensations	.67		
6. Even when something is important to me, I'll rarely do it if there is a chance it will upset me	.56		
7. I work hard to keep out upsetting feelings	.78		
8. When unpleasant memories come to me, I try to put them out of my mind	.56		
9. I can experience my thoughts and feelings as they are, without trying to change them*	.62		
10. I can take thoughts and feelings as they come, without attempting to control or avoid them*	.76		
11. I am willing to fully experience whatever thoughts, feelings and sensations come up for me, without trying to change or defend against them*	.72		
12. I get so caught up in my thoughts that I am unable to do the things that I most want to do	.46		
13. I don't allow thoughts to interfere with what I want to do*	.50		
14. Thoughts are just thoughts – they don't control what I do*	.57		
15. It seems I am "running on automatic" without much awareness of what I'm doing			.64
16. Even when doing the things that matter to me, I find myself doing them without paying attention			.89
17. I rush through meaningful activities without being really attentive to them			.77
18. I do jobs or tasks automatically, without being aware of what I'm doing			.74
19. I find it difficult to stay focused on what's happening in the present			.53

20. I have values that give my life more meaning*	.56
21. I have important values I choose to live by*	.69
22. I make choices based on what is important to me, even if it is stressful*	.70
23. My values are really reflected in my behaviour*	.76
24. I am able to follow my long terms plans including times when progress is slow*	.69
25. I can keep going with something when it's important to me*	.70
26. I behave in line with my personal values*	.75
27. I undertake things that are meaningful to me, even when I find it hard to do so*	.73
28. I act in ways that are consistent with how I wish to live my life*	.70
29. I can identify the things that really matter to me in life and pursue them*	.74
30. I know what matters to me and where I want to end up in life*	.59

Note. *items are positively valenced and were reversed scored.

3.5. Reliability of the CompACT

Cronbach alpha values are liable to being artificially inflated, either by having a large number of items, or by having items which are highly inter-correlated (Cortina, 1993). Clark and Watson (1995) therefore advocate calculating an average of the interitem-total correlations amongst a measure's items to determine a measure's internal consistency. The CompACT's average item-total correlation (.31) based on 30 items and 352 cases fell within the recommended ranges for adequate internally consistency (i.e. .15-.50; see Briggs & Cheek, 1986).

3.6. Validity of the CompACT

Table 5 below summarises the associations between the CompACT and its three subscales with other measures.

3.6.1. Convergent validity. The CompACT demonstrated a large significant correlation[4] with the AAQ-II (r = .78), thus indicating good convergent validity with established measures of ACT processes[5]. The CompACT's

[4] The magnitude of correlations are based on Cohen's (2013) classification: .10 (small); .30 (medium); and >.50 (large).

[5] The final version of CompACT did not include any AAQ-II items and thus, the measure's correlation with the AAQ-II is not inflated by possessing shared items with this measure.

'openness' subscale had the strongest association with the AAQ-II (*r* = .77), followed by the 'self-awareness' subscale (*r* = .54) and the 'motivation-activation' subscale (*r* = .38).

3.6.2. Discriminant validity. No significant correlation was observed between either the CompACT and the short-form MCSD, and the AAQ-II and the short-form MCSD.

3.6.3. Concurrent validity. The CompACT was found to have large positive correlations (*rs* = .56-.64) with all three subscales representing the DASS-21, and correlated most strongly with the depression subscale (*r* = .64) (See Table 5). In comparison, the AAQ-II exhibited larger positive correlations than the CompACT (*rs* = .64-.72) with all three subscales of the DASS-21. Like the CompACT, the AAQ-II also correlated most strongly with the DASS-21 depression subscale (*r* = .72). When the CompACT's three subscales were examined for their associations with the DASS-21 subscales, the openness subscale (*rs* = .53-55) correlated more strongly with all of the DASS-21 subscales compared to the CompACT's self-awareness (*rs* = .43-.48) and motivation-activation subscales (*rs* = .20-.43).

The CompACT demonstrated a large significant negative correlation with the SF-12v2's mental health subscale (*r* = -.65), and a small but significant negative correlation with its physical health subscale (*r* = -.23). Overall, the CompACT demonstrated a similar order of magnitude of association with the SF-12v2 compared to the AAQ-II on both mental health and physical health subscales, with the AAQ-II having a slightly higher correlation with the SF-12v2 mental health subscale (*r*= -.72).

The CompACT's openness subscale had the strongest correlation with the SF-12v2's mental health subscale (*r* = -.58). All three of the CompACT's subscales mirrored the CompACT's stronger associations with the SF-12v2 mental health subscale (*rs* = -.37-.58) compared to the physical health subscale (*rs* = .15-.19).

Table 5

Correlations Between the CompACT, the CompACT's Subscales, the AAQ-II and Other Measures (N= 352)

	Correlation (*r*)				
Measure	**CompACT total score**	**CompACT openness subscale**	**CompACT self-awareness subscale**	**CompACT motivation-activation subscale**	**AAQ-II**
AAQ-II	.78*	.77*	.50*	.38*	-
DASS-21:					
Depression	.64*	.55*	.48*	.43*	.72*
Anxiety	.56*	.53*	.43*	.28*	.66*
Stress	.56*	.55*	.45*	.20*	.66*
SF-12v2					
Physical health Subscale:	-.23*	-.19*	-.19*	-.15*	-.23*
General health	.04	.01	.09	.02	.01
Physical functioning	-.13*	-.07	-.18	-.10	-.10*
Role physical	-.28*	-.23*	-.25*	-.17*	-.27*
Bodily pain	-.27*	-.25*	-.12*	-.17*	-.28*
Mental health subscale:	-.65*:	-.58*	-.51*	-.37*	-.72*
Role emotion	-.58*	-.58*	-.41*	-.35*	-.65*
Mental health	-.61*	-.56*	-.68*	-.32*	-.68*
Vitality	-.44*	-.36*	-.40*	-.27*	-.44*
Social Functioning	-.54*	-.49*	-.41*	-.31*	-.62*
Short-form MCSD	.01	.01	-.01	.01	-.01

Abbreviations: AAQ-II (Acceptance and Action Questionnaire, version 2); DASS-21 (Depression Anxiety and Stress Scales, 21-item version); SF-12v2 (Short-form Health and Survey, 12-item version); MCSD (Marlowe Crowne Social Desirability Scale, 12-item version).

Note. *Correlation is significant at $p < .01$

4. Discussion

Current measures of ACT processes are limited, particularly in relation to possessing items which either: 1) overlap with distress constructs that are theoretically distinct from ACT processes; and/or 2) which operationalise/measure ACT processes in isolation from other ACT processes, and thus fail to provide a comprehensive account of behavioural functioning. There are also questions as to whether ACT's conceptual model underpinning ACT process measures is comprised of fewer than six distinct processes. The aims of this study were to develop a more representative general measure of ACT processes and test the psychometric viability of Hayes et al.'s (2011) three dyadic ACT processes as the conceptual framework for the CompACT.

In phase one a Delphi consensus methodology was used to select 37 items considered to have the best face and content validity by ACT experts. In phase two, participants completed the CompACT in addition to measures other measures to assess different aspects of the CompACT's validity. A stable three-factor solution was found for a 30-itemed version of the CompACT. Adequate average item-total correlations between the CompACT's items, and high Cronbach's alpha coefficients for the CompACT's three subscales attested to the measure's overall reliability and internal consistency. Significantly, an examination of the content of items loading onto the CompACT's three factors indicated thatevery item loaded onto the specific factor it was hypothesized to. There is therefore both conceptual and statistical grounds supporting that each of the CompACT's items is an accurate representation of the particular ACT process it was designed to measure. Additionally, findings suggest higher scores on the CompACT were associated with: higher scores on the AAQ-II; higher scores on the DASS-21; and lower scores on the SF-12V2. The CompACT therefore performed in theoretically predictable ways by supporting ACT's central hypothesis: higher levels of psychological inflexibility correlate with higher levels of psychological distress and lower levels of health and general wellbeing. Findings also suggest that the CompACT possesses adequate divergent validity, given its lack of association with a measure of socially desirable responding.

The Delphi methodology ensured that the CompACT's items were considered to have good face and content validity by those with expertise in ACT. Significantly, all items taken from the AAQ-II failed to reach consensus thresholds for inclusion in the CompACT. This further supports Wolgast's (2014) and our criticisms of the AAQ-II's items, and suggests that the CompACT possesses items which more precisely represent ACT processes based upon expert judgement. A Delphi methodology was also advantageous by enabling ACT experts to rate items anonymously. Panel members were therefore arguably less susceptible to the kinds of social psychological processes known to influence and shape group decision-making (see Turner, Brown & Tajfel, 1979), which could have affected the development of the AAQ-II and other ACT process measures which did not use this methodology (see extended paper 4.1.).

The CompACT possesses both positively and negatively valenced items, and this was determined by several factors including: 1) how authors of the ACT process measures I extracted items from connoted items either positively or negatively; 2) how experts rated items during the Delphi study resulting in either them being retained/excluded; and 3) how each item contributed to the CompACT's factor structure resulting in them being retained/excluded (see extended paper 4.2). Research suggests that incorporating items of opposing valence within a single measure can undermine the resolution of a stable factor structure, specifically because negatively valenced items can load onto a separate factor unrelated to the construct(s) of interest due to sharing a common variance (DiStefano & Motl, 2006; Marsh, 1996; Roszkowski & Soven, 2010). Although the CompACT separated into factors with entirely negatively or positively valenced items, the factors identified are consistent with our expectation that different items would capture different constructs; all negatively valenced items represent openness and self-awareness dyadic processes, and all positively valenced items represent the motivation-activation dyadic process. The CompACT's three factor structure is therefore unlikely to be an artefact of item valence/phrasing, and instead represents 'real' theorised latent constructs that are conceptually distinct.

It is of note that the purported method effect of negatively valenced items is not universally agreed upon, with some suggesting (as per our findings) that negatively valenced items may actually reflect true underlying construct(s) (Lance, Baranik, Lau & Sharlau, 2009). Moreover, despite being a mixed valenced measure, I was still able to find a stable three factor structure for the CompACT. This suggests that including items with a mixed valence may not artificially affect a measure's factor structure per se (see extended paper 4.3).

Both the CompACT and the AAQ-II infer psychological flexibility by reverse scoring negatively valenced items. Kashdan and Rottenberg (2010) argue however, that psychological flexibility and inflexibility are not merely polar opposite constructs of each other; instead, they suggest that psychological flexibility is on a continuum, with the extreme end point of the flexibility continuum being inflexibility. Based on this conceptualisation, reverse scoring items with extreme responses may pose a problem, as it is questionable whether the presence of one extreme of the continuum can be inferred from negation of the opposite extreme. Despite this potential issue, many of the ACT process measures I used to extract items from adopt a similar practice of operationalising items in terms of the 'absence' of a particular ACT process, and then reverse scoring such items to derive the 'presence' of that ACT process. It may be therefore be the case that psychological flexibility is difficult to define, and is more readily operationalised in terms of its absence.

To our knowledge, I amthe first to explore and provide support for Hayes et al.'s (2011) conceptualisation of ACT's clinical model as theoretically and psychometrically viable. This finding also more broadly strengthens the argument that ACT's conceptual model may be composed of fewer than six distinct processes. Pragmatically, having fewer processes to target within ACT treatment may have clinical utility. Indeed, some have already simplified ACT's conceptual model as a 'triflex' with a similar amalgamation of ACT processes to Hayes et al.'s model, in order to make implementing ACT interventions more manageable (see Harris, 2009).

An examination of the intercorrelations between factors (*rs* = .25-.39) suggests that there are small and significant correlations between factors. Given that no

factor correlated above .5 this is further in accordance with ACT theory that processes may be interrelated, but also distinct (see extended paper 4.4.). The CompACT's openness subscale had the strongest association with the AAQ-II. This finding is perhaps not surprising given our earlier suggestion that the AAQ-II's items primarily reflect acceptance/non-acceptance and defusion/fusion processes within the hexaflex (which are represented in our measure by the openness subscale). The small to moderate correlation between the CompACT's motivation-activation subscale and the AAQ-II further substantiates our suggestion that the AAQ-II does not capture other important ACT processes, such as values and committed action (see extended paper 4.5.). This criticism was echoed by ACT experts within our study, and resulted in the AAQ-II's items failing to achieve sufficient ratings for inclusion within the CompACT.

The CompACT's correlation with the DASS-21 is not so high as to suggest that these measures are assessing the same construct (Nunnally & Bernstein, 1994). The CompACT was also found to have weaker associations with all measures of distress within the DASS-21 (and in particular anxiety and stress subscales) compared to the AAQ-II. This finding, coupled with the CompACT's higher face and content validity ratings compared to the AAQ-II therefore implies that the CompACT is less likely to be a conflated distress measure (see extended paper 4.6.). Gámez et al. (2011) suggest that participants unfamiliar with the ACT model may have difficulty in distinguishing ACT processes from symptoms of psychological distress as operationalised in self-report measures like the DASS-21. If this was the case in our study, the high correlation between the CompACT and DASS-21 may be an artefact of participants conflating these two measures due to naivety to the ACT model. Further research around this issue is therefore warranted to rule this out (see extended paper 4.7.).

The correlation between the CompACT and the SF-12v2 is also not so high as to suggest that the CompACT is actually a health and wellbeing measure. Despite the dearth of CompACT items which reflect a physical health component, there was, a small but significant association between the CompACT and the SF-12v2's physical health subscale. This association

provides further evidence that the CompACT is measuring something that is theoretically distinct from health and wellbeing (i.e. psychological inflexibility) and is inversely related to this outcome (i.e. higher levels of psychological inflexibility are associated with lower levels of health and wellbeing - see extended paper 4.8.).

Overall then, the CompACT represents an alternative comprehensive single measure of multiple ACT processes to the AAQ-II, whilst also benefitting from possessing items with better face and content validity than this measure. Although I suggest that the AAQ-II would probably account for more of the variance in distress outcome measures than the CompACT, I would argue that this would be an artefact of the conflation between these two measures, rather than evidence attesting to the AAQ-II's better predictive validity. Thus, it may be useful for future research to compare and evaluate these measures in terms of accounting for personally meaningful functional outcomes and committed activity.

The clinical utility of the CompACT over and above the AAQ-II is enhanced by its incorporation of three process subscales, as well as an overall measure of psychological flexibility. This renders it suitable for use in process research into ACT's mechanisms of change such as component studies, and/or single case experimental designs (SCEDs) research which have a strong tradition of being used to inform and develop psychological theory within a behavioural paradigm (Smith, 2012 - see extended paper 4.9.). Equally, the CompACT could also have utility in clinical settings as a means to capture and monitor the proposed mediators of change within the ACT model, and thereby support ACT's conceptualisation of psychological distress over and above extant therapies. Mediational analytic research has been advocated as an effective means by which to compare components of ACT and CBT, and then see if any changes accord more with ACT's theorised processes of change (Lundgren, Dahl & Hayes, 2008).

However, there are several limitations to our study which warrant discussion. There were few items representing the self as context ACT process within the initial item pool. This was mainly due to the fact that only one measure of self as

context was found during our literature search, and this had not undergone extensive psychometric validation. Some items from this measure were later added to our item pool in addition to our own items and items supplied from the expert panel representing this process. However, none of these items attained sufficient ratings to be included in the CompACT. Some panel members justified their low ratings of such items, suggesting that self as context was defined in too much of an abstract way to be comprehensible. ACT interventions typically utilise metaphor exercises in order to connote the abstract principles of self as context and render them more meaningful (see Harris, 2009). Whilst socialisation to the ACT model may therefore help people respond to self as context items, participants completing the measure in our study were from a non-clinical sample and were unlikely to have undergone ACT interventions. I therefore prioritised having items that were comprehensible, even if that meant the self as context as process was not represented within our item set.

More broadly however, this issue raises questions as to whether 'static' self-report measures are in and of themselves sufficient and complementary to the 'dynamic' ACT processes they are attempting to capture (Kashdan & Rottenberg, 2010). Hayes, Barnes-Holmes and Wilson (2012) suggest that because ACT interventions can undermine the social/verbal contexts that maintain the consistent responding that psychometric methods aim to analyse, it may the case that alternative measures of ACT's processes of change, such as the Implicit Relational Assessment Procedure (IRAP; see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2011) are more suitable.

The CompACT possesses an uneven distribution of items representing different ACT processes, with defusion/fusion having the lowest number of items. A potential issue with this is that the breadth and scope of a particular ACT process may not be adequately represented. This uneven distribution was the product of a number of factors, including: 1) an initial uneven distribution of items representing different ACT processes within the preliminary 106 item pool, due to a higher prevalence of measures of certain ACT processes over and above others processes within the extant literature; 2) how ACT experts rated the face and content validity of items, resulting in them being either

retained/excluded; and 3) the removal of certain items which comprised the resolution of the CompACT's factor structure. Whilst I could have included additional defusion items for further rating in subsequent rounds, consensus had already been achieved for the existing items, and further rating could have led to burden and therefore attrition of ACT experts from our panel.

Finally, ACT theory suggests that psychological flexibility mediates distress and wellbeing. However, cross-sectional designs, such as those used within our study are insufficient to demonstrate temporal mediation. This lends further credence to the suggestion that additional behavioural measures are needed in order to attest to the mediating effect of ACT processes on key outcomes (see Kashdan & Rottenberg, 2010).

4.1. Future research

The CompACT's factor structure, and its validity and reliability require additional testing and cross validation to further substantiate its psychometric robustness. Based on the test construction literature (e.g. DeVellis, 1991) this research could proceed via the following stages: 1) test the CompACT's factor structure a priori through a confirmatory factor analysis (CFA); 2) compare how participants perform on the CompACT from clinical versus non-clinical settings to see if the measure can distinguish between known groups and thus further demonstrate its construct validity; 3) use item factor loading data to help reduce the number of items within the CompACT, and thereby make it a more brief measure that has better utility in applied settings (see Lindell & Whitney, 2001 - see extended paper 4.10); and 4) establish the CompACT's suitability for repeated administration (i.e. test-retest reliability) which is a particularly important component of any ACT measure, given that ACT processes are purported to be dynamic and changeable with context.

5. Conclusion

The CompACT shows initial promise as a comprehensive measure of Hayes et al.'s (2011) three dyadic ACT processes. but requires further research to support it as a psychometrically robust alternative to the AAQ-II.

Journal paper word count: 6000

References

Ballard, R. (1992). Short forms of the Marlowe-Crowne social desirability scale. *Psychological Reports*, *71*(3), 1155-1160.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2011). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record*, *60*(3), 527-542.

Blackledge, J. T., & Drake, C. E. (2013). Acceptance and Commitment Therapy: Empirical and theoretical considerations. In S. Dymond, & B. Roche (Ed.), *Advances in relational frame theory: Research and application* (pp. 219-252). Oakland: Context Press.

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., ... & Zettle, R. D. (2011). Preliminary psychometric properties of the acceptance and action questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, *42*(4), 676-688.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, *54*(1), 106-148.

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, *84*(4), 822-848.

Cardaciotto, L., Herbert, J. D., Forman, E. M., Moitra, E., & Farrow, V. (2008). The assessment of present-moment awareness and acceptance the Philadelphia mindfulness scale. *Assessment*, *15*(2), 204-223.

Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Multivariate Behavioral Research*, *12*(3), 289-325.

Chawla, N., & Ostafin, B. (2007). Experiential avoidance as a functional dimensional approach to psychopathology: An empirical review. *Journal of Clinical Psychology*, *63*(9), 871-890.

Ciarrochi, J. Bilich, L., & Godsel, C. (2010). Psychological flexibility as a mechanism of change in acceptance and commitment therapy. In Ruth Baer's (Ed), *Assessing mindfulness and acceptance: Illuminating the processes of change*.(pp. 51-76). Oakland, CA: New Harbinger Publications.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological assessment*, *7*(3), 309-319.

Cohen, J. (2013). *Statistical power for the behavioural sciences* (rev. Ed). New York, NY: Academic Press.

Cohen, R. J., & Swerdlik, M. E. (2005). Psychological testing and measurement: An introduction to tests and measurement. New York, NY: McGraw-Hill.

DeVellis, R.F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage Publications.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401-409.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, *13*(3), 440-464.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London, UK: Sage.

Forman, E. M., Herbert, J. D., Juarascio, A. S., Yeomans, P. D., Zebell, J. A., Goetter, E. M., & Moitra, E. (2012). The Drexel Defusion Scale: A new measure of experiential distancing. *Journal of Contextual Behavioral Science*, *1*(1), 55-65.

Gámez, W., Chmielewski, M., Kotov, R., Ruggero, C., Suzuki, N., & Watson, D. (2014). The Brief experiential avoidance questionnaire: Development and initial validation. *Psychological assessment*, *26*(1), 35-45.

Gámez, W., Chmielewski, M., Kotov, R., Ruggero, C., & Watson, D. (2011). Development of a measure of experiential avoidance: The multidimensional experiential avoidance questionnaire. *Psychological Assessment*, *23*(3), 692-713.

Gillanders, D. T., Bolderston, H., Bond, F. W., Dempster, M., Flaxman, P. E., Campbell, L., ... & Remington, B. (2014). The development and initial validation of the cognitive fusion questionnaire. *Behavior Therapy*, *45*(1), 83-101.

Gird, S. R., & Zettle, R. (2013). *Developing a quantitative measure of self-as-context*. In Proceedings: 9th Annual Symposium: Graduate Research and Scholarly Projects. Wichita, KS: Wichita State University, p.47-48

Greco, L. A., Lambert, W., & Baer, R. A. (2008). Psychological inflexibility in childhood and adolescence: Development and evaluation of the Avoidance and Fusion Questionnaire for Youth. *Psychological Assessment, 20(2)*, 93-102.

Gregg, J. A., Callaghan, G. M., Hayes, S. C., & Glenn-Lawson, J. L. (2007). Improving diabetes self-management through acceptance, mindfulness, and values: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *75*(2), 336-343.

Harris, R. (2009). *ACT made simple: An easy-to-read primer on acceptance and commitment therapy*. Oakland, CA: New Harbinger Publications.

Hayes, S. C. (2004). Acceptance and commitment therapy and the new behavior therapies: Mindfulness, acceptance, and relationship. In V. M. F. S. C. Hayes, & M. M. Linehan (Ed.), *Mindfulness and acceptance: Expanding the cognitive-behavioral tradition* (pp. 1-29). New York, NY: Guilford Press.

Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual behavioral science: Creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioral Science*, *1*(1), 1-16.

Hayes, S. C., Levin, M. E., Plumb-Vilardaga, J., Villatte, J. L., & Pistorello, J. (2013). Acceptance and commitment therapy and contextual behavioral science: Examining the progress of a distinctive model of behavioral and cognitive therapy. *Behavior Therapy*, *44*(2), 180-198.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, *44*(1), 1-25.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. New York, NY: Guilford Press.

Hayes, S. C., Villatte, M., Levin, M., & Hildebrandt, M. (2011). Open, aware, and active: Contextual approaches as an emerging trend in the behavioral and cognitive therapies. *Annual Review of Clinical Psychology*, *7*(4), 141-168.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *44*(2), 227-239.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393-416.

Herzberg, K. N., Sheppard, S. C., Forsyth, J. P., Credé, M., Earleywine, M., & Eifert, G. H. (2012). The Believability of Anxious Feelings and Thoughts Questionnaire (BAFT): A psychometric evaluation of cognitive fusion in a nonclinical and highly anxious community sample. *Psychological assessment*, *24*(4), 877.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179-185.

Hsu, C.C., & Sanford, B.A. (2007). The delphi technique: Making sense of consensus. *Practical Assessment, Research and Evaluation*, *10*(12), 1-8.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401-415.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31-36.

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, *30*(7), 865-878.

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, *3*, 1–27.

Lance, C. E., Baranik, L. E., Lau, A. R., & Scharlau, E. A. (2009). If it ain't trait it must be method: (Mis)application of the multi trait multi method, In C. E. Lance & R. J. Vandenberg (Eds.), Statistical and methodological myths and urban legends: Received doctrine, verity and fable in organisation and social research (pp.339-362). New York: Routledge.

Lang, A. J., Schnurr, P. P., Jain, S., Raman, R., Walser, R., Bolton, E., ... & Benedek, D. (2012). Evaluating transdiagnostic treatment for distress and impairment in veterans: A multi-site randomized controlled trial of acceptance and commitment therapy. *Contemporary Clinical Trials*, *33*(1), 116-123.

Levin, M., & Hayes, S. C. (2009). Is acceptance and commitment therapy superior to established treatment comparisons? *Psychotherapy and Psychosomatics, 78*(6)*,* 380-381.

Lillis, J., & Hayes, S. C. (2008). Measuring avoidance and inflexibility in weight related problems. *International Journal of Behavioral Consultation and Therapy*,*4*(4), 348-354.

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, *86*(1), 114-121.

Lundgren, T., Dahl, J., & Hayes, S. C. (2008). Evaluation of mediators of change in the treatment of epilepsy with acceptance and commitment therapy. *Journal of Behavioral Medicine*, *31*(3), 225-235.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors?. *Journal of Personality and Social Psychology*, *70*(4), 810-819.

McCracken, L. M., Chilcot, J., & Norton, S. (2015). Further development in the assessment of psychological flexibility: A shortened Committed Action Questionnaire (CAQ-8). *European Journal of Pain*, *19*(5), 677-685.

Netemeyer, R.G., Bearden, W.O., & Sharma, S. (2003) *Scaling procedures: Issues and applications.* Thousand Oaks, CA: Sage Publications.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory (3rd ed.)*. New York, NY: McGraw-Hill.

Öst, L. G. (2008). Efficacy of the third wave of behavioral therapies: A systematic review and meta-analysis. *Behaviour Research and Therapy*, *46*(3), 296-321.

Öst, L. G. (2014). The efficacy of Acceptance and Commitment Therapy: An updated systematic review and meta-analysis. *Behaviour Research and Therapy*, *61*(10), 105-121.

Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, *35*(1), 113-130.

Ruiz, F. J., Langer Herrera, A. I., Luciano, C., Cangas, A. . & Beltran, I. (2013). Measuring experiential avoidance and psychological inflexibility; The Spanish version of the Acceptance and Action Questionnaire - II. *Psicothema, 25*(1), 123-129.

Singh, P., Aggarwal, R., Zevin, B., Grantcharov, T., & Darzi, A. (2014). A global Delphi consensus study on defining and measuring quality in surgical training. *Journal of The American College of Surgeons*, *219*(3), 346-353.

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, *17*(4), 510-550.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99-103.

Tabachnick, B.G. and Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). New York, NY: Allyn and Bacon.

Trompetter, H. R., Ten Klooster, P. M., Schreurs, K. M., Fledderus, M., Westerhof, G. J., & Bohlmeijer, E. T. (2013). Measuring values and committed action with the Engaged Living Scale (ELS): Psychometric evaluation in a nonclinical sample and a chronic pain sample. *Psychological Assessment*, *25*(4), 1235-1246.

Turner, J. C., Brown, R. J., & Tajfel, H. (1979). Social comparison and group interest in ingroup favouritism. *European Journal of Social Psychology*, *9*(2), 187-204.

Vowles, K. E., McCracken, L. M., McLeod, C., & Eccleston, C. (2008). The chronic pain acceptance questionnaire: Confirmatory factor analysis and identification of patient subgroups. *Pain*, *140*(2), 284-291.

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*(3), 220-233.

Ware, J. E., Kosinski, M., Turner-Bowker, D. M., & Gandek, B. (2002). User's manual for the SF-12v2 Health Survey. *Lincoln, RI: QualityMetric Incorporated*.

Wilson, K. G., Sandoz, E. K., Kitchens, J., & Roberts, M. (2011). The valued living questionnaire: Defining and measuring valued action within a behavioral framework. *The Psychological Record*, *60*(2), 249-272.

Wolgast, M. (2014). What does the acceptance and action questionnaire (AAQ-II) really measure? *Behavior Therapy*, *45*(6), 831-839.

Zevin, B., Bonrath, E. M., Aggarwal, R., Dedy, N. J., Ahmed, N., & Grantcharov, T. P. (2013). Development, feasibility, validity, and reliability of a scale for objective assessment of operative performance in laparoscopic gastric bypass surgery. *Journal of the American College of Surgeons*, *216*(5), 955-965.

EXTENDED PAPER

1. Extended introduction

1.1. The development of ACT

ACT is a relatively new therapy that emerged during the late 1990's, but it can be traced back further to its earliest form as 'cognitive decentring' during the 1980's (see Hayes & Brownstein, 1986). ACT principally developed out of research showing that behavioural principles (circa 1957-1986) provided an insufficient account of the role of language and cognition in the development of clinical problems. More broadly, theses limitations, combined with dissatisfaction amongst the research community of behaviourism's treatment of animal behaviour as an analogue of human behaviour (e.g. White, Juhasz & Wilson, 1973), led to the previously dominant behavioural approaches being usurped by more cognitive therapies in a 'cognitive revolution' (Blackledge & Drake, 2013). Hayes (2004a) notes however, that whilst such cognitive therapies were proving clinically useful in the short-term, their development from clinical theories divorced from experimental and applied research posed longer term problems. Indeed, some argue that this 'top-down' approach to understanding clinical phenomena has limited our ability to elaborate on what the mechanisms of change are that underpin cognitive therapy (e.g. Hofmann & Asmundson, 2008), which some have suggested still remain unclear (Longmore & Worrell, 2007).

In a bid to examine the role of language and cognition in clinical problems, buttressed by behaviourism's basic theoretical tenets and tradition of lab-based investigation, Hayes and colleagues set about developing a behavioural account of human language and cognition. From this research emerged relational frame theory (RFT; Hayes, Strosahl, Barnes-Holmes & Roche, 2001), and later then ACT as a clinical representation and application of RFT. RFT is a broad and complex theory that requires consideration of many of its technical aspects in order to fully grasp it. For the purposes of this thesis however, only those aspects of RFT which most clearly relate to ACT and its development are considered (for an extended discussion of RFT see Hayes et al., 2001).

At its core, RFT treats language and cognition synonymously[6] as behaviour that is learned over time, and is subject to the same contingencies of reinforcement and punishment as other forms of behaviour (Törneke, 2010). Whilst this notion does not deviate from Skinner's (1957) much criticised thesis on verbal behaviour, RFT extends this by asserting that language is a special class of 'relational operant' behaviour, that enables humans to relate any stimuli to each other based not only on their physical properties (e.g. size, weight, etc.), but also in terms of their non-physical or arbitrary properties (e.g. 'tenderness', 'malevolence', 'sincerity' etc.) (Dymond & Rehfeldt, 2000). Being able to relate stimuli in this latter way is referred to as arbitrarily applicable relational responding (AARR). Where stimuli are related to each other through AARR, their properties or function can change depending on the context; this is referred to as transformation of stimulus function. Based on these processes of AARR and transformation of stimulus function, stimuli can come to acquire properties that they never previously had, simply through participating in a verbal relationship with other stimuli. More crucially, as a consequence of deriving stimulus relations, the ways in which I respond to such stimuli (overtly or covertly) can also change.

As an example of the above process, consider a person who experiences being trapped in a car crash. The 'car' may come to have aversive properties through normal classical conditioning processes (i.e. it becomes associated with being trapped and in danger based on a learning history) and produces an anxiety-fear response, potentially resulting in avoidance of cars in the future. According to RFT however, the word 'trapped' may also come to acquire an aversive function, as it participates in a verbal relationship with the actual experience of being trapped. This may lead to a person overtly avoid talking about being 'trapped', but also covertly and 'experientially' avoiding this (e.g. through suppressing thoughts and/or memories relating to being trapped). Moreover, as the concept of 'being trapped' is now aversive, situations in the future which may be associated (or more precisely participate in a verbal relationship of

[6] Language, cognition and thinking can be considered as interchangeable in RFT, as they involve the use of overt and covert symbols (Hayes et al., 2001).

'coordination'[7]) with 'being trapped' (e.g. trapped in a relationship) could also produce the same anxiety-fear response and subsequent behavioural and experiential avoidance.

The above example demonstrates that, whilst these basic relational processes underpinning language and cognition may have given the human species a considerable advantage in our own survival[8], the ability to derive stimulus relations has broadened our interface with aversive events and therefore our experience of emotional suffering (Hayes et al., 1999); we can summon painful memories from the past to bear on the present and re-experience this pain, despite no immediate environmental cues; we can make predictions about aversive consequences yet to happen, and experience fear even if we are safe in the present; and we can compare and contrast ourselves to others on the basis of seemingly apparent immutable 'facts' (e.g. our 'badness') without seeing how such facts are constructed through language (Flaxman et al., 2010). It is this central idea within RFT that language processes can begin to override our direct experience of the world as it is, which provides the theoretical underpinning to ACT's theory of psychopathology.

1.2. RFT and ACT's theory of psychopathology

ACT adopts RFT's concepts of AARR and transformation of stimulus function to describe how previously neutral internal private events such as thoughts, feelings, memories and sensations can come to acquire aversive functions and dominate overt behaviour. In ACT parlance, this process is called cognitive fusion, and is considered to relate to the failure to notice the process of thinking (context) over the products of thinking (content) (Hayes, Luoma, Bond, Masuda, & Lillis, 2006). For example, the consequence of fusion with one's thoughts means that such thoughts are taken literally and considered immutable truths, rather than being seen for what they actually are; transient events which have verbal content with no objective 'truth'. Cognitive fusion promotes behaviour

[7] There are different ways in which stimuli can be verbally related to each other including; opposingly (i.e. in terms of difference); deictically (i.e. in relation to perspective); causally (i.e. x caused y to break) etc.

[8] Relation framing would have allowed us to avoid aversive events (e.g. certain areas with predators) without ever having needed to have experienced them before, simply by hearing someone tell us 'don't go there, it is dangerous' or some simila form of verbal communication. In order to understand how a verbal 'instruction' from another person may influence behaviour in the absence of previous experience of an event, see the discussion on rule governed behaviour (RGB) 1.4.4.

which is predominantly shaped by such thoughts, rather than by the direct contingencies of a person's external world. Whilst cognitive fusion in the main is an advantageous process (e.g. if one hears 'watch out for that car', cognitive fusion permits us to take decisive action without having to directly see the car; Hayes, Pistorello & Levin, 2012), chronic cognitive fusion can lead to a less values-consistent way of living.

ACT further suggests that cognitive fusion provides the foundation for the development of clinical problems when one attempts to 'experientially avoid' the aversive internal events one has become 'cognitively fused' and entangled with, as inevitably this results in an additional loss of contact with positive reinforcing events that give an individual's life meaning (Hayes et al., 1999). Whilst avoidance of externally located aversive stimuli may be possible in the long-term, the same is not true for internally located aversive stimuli. Moreover, evidence suggests that attempts to avoid and suppress aversive internal experiences (e.g. avoiding intrusive thoughts through distraction or suppression) can paradoxically amplify their intensity, frequency and duration[9] (e.g. Wenzlaff & Wegner, 2000). ACT therefore suggests that it is one's responses to internal experiences rather than the content of such experiences which determines psychological well-being, with avoidance garnering greater psychological distress than acceptance.

1.3. The clinical focus of ACT

Hayes (2004a) describes behavioural therapies as having undergone three broad iterations, or 'waves': 'first wave' behavioural therapies used basic principles of classical and operant conditioning to describe the development and maintenance of clinical problems; 'second wave' behavioural therapies embraced cognitive theories tenets on the role of thoughts in the development of psychological distress, and focussed on subjecting these thoughts to testing and disputation in order to exchange the content of thoughts for something more closely aligned with reality, or provide an alternative thought that results in

[9] RFT describes this finding in terms of relational networks amongst stimuli, which can be strengthened and embellished through cognitive avoidance, by bringing further stimuli into the relational frame of the stimulus being avoided; thus the list of stimuli which serve as 'reminders' of the original feared stimulus continues to expand and repeatedly produce the same fear-response (Wilson, Hayes, Gregg & Zettle, 2001).

less emotional distress; 'third wave' approaches deviated from targeting the form, frequency or situational sensitivity of aversive private events, to targeting their function and analysing the context in which they are triggered and acquire their aversive functions. This shift in focus developed due to a number of different critiques of the second wave approaches on both theoretical and empirical grounds (see Hayes, 2004a for a further discussion of this). Other third wave approaches alongside ACT include dialectical behavioural therapy (DBT; Linehan, 1993), compassion focussed therapy (CFT; Gilbert, 2009) and mindfulness-based cognitive behavioural therapy (MB-CBT; Segal, Williams, & Teasdale, 2002).

1.4. ACT's hexaflex model and the six core processes

The hexaflex (see Figure 1 below) is considered both a model of functional behaviour, and a model of psychopathology, as ACT theory proposes that psychopathology is grounded in healthy rather than dysfunctional psychological processes (Hayes et al., 2012). Indeed, each of the six ACT processes within the hexaflex can be inverted to describe ACT's conceptualisation of psychopathology (Hayes et al., 2006). Before going on to describe each of the six ACT processes within the hexaflex in more detail, it is worth noting that the hexaflex model represents Hayes and colleagues' attempts to translate the complex verbal relational processes found within RFT lab-based research, into more accessible and clinically focussed 'mid-level' concepts[10]. As such, ACT's mid-level concepts are not behaviour-analytically defined, and therefore do not provide the same level of theoretical precision as their underpinning RFT counterparts (to understand how this may be an issue, see journal article 1.2.3). For the purposes of this thesis however, the following discussion of the six core processes in the hexaflex is tied to ACT's mid-level operational definitions of RFT concepts.

[10] In order to better explain this translation process, a metaphor of a computer system can be used. Mid-level concepts are akin to the software of a computer than is readily accessible by users, with such software being underpinned by the binary coding of computer programming which may represent RFT concepts. Crucially, just as one does not need to understand computer programming in order to use such software, one does not necessarily have to understand RFT in order to apply the mid-level concepts of ACT which derive from RFT.

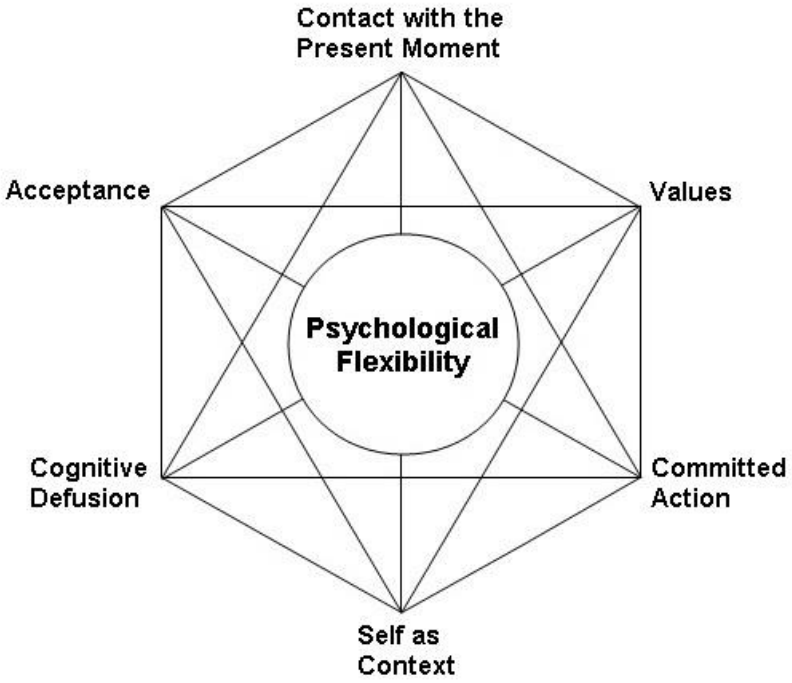


Figure 1. ACT's hexaflex model (adapted from Hayes et al., 1999).

1.4.1. Acceptance - experiential avoidance. Acceptance is the willingness to actively embrace private events that may be aversive, without trying to change their frequency or form (Hayes et al., 1999). Acceptance is not an end within itself but is the necessary step that enables commitment to values-based action. This point is crucial to make, as acceptance without any connection to values based action may be connoted with 'tolerance', 'soldiering on' or 'grinning and bearing it'; all of which are likely to exacerbate distress through producing behaviour that is under aversive rather than appetite control (Hayes et al., 2006). It is also important to note that acceptance is an active process involving a deliberate and curious exploration of one's internal private events such as thoughts, feelings, memories and sensations (Harris, 2007). Such exploration not only opens up the opportunity to see these internal experiences for what they are rather than what our mind says they are (i.e. there is a 'de-literalisation' process; see 1.4.2.), but also permits one to begin to respond on the basis of direct contingencies and experiences; in other words, acceptance promotes the (psychological) flexibility and freedom to engage with

those aspects of life which are meaningful, without being inhibited by aversive internal experiences (e.g. self-critical thoughts).

Acceptance is a central component of ACT treatment and has been found to mediate outcomes in a number of areas including: chronic pain (Åkerblom, Perrin, Rivano, & McCracken, 2015); anxiety and depression (Forman, Herbert, Moitra, Yeomans & Geller, 2007); quality of life in patients with seizures (Lundgren, Dahl & Hayes, 2008); work-place stress (Bond & Bunce, 2003); smoking cessation (Gifford et al., 2004); and diabetes management (Gregg, Callaghan, Hayes, & Glenn-Lawson, 2007). However, there has been criticism of the measures used to assess acceptance within these studies, particularly when the AAQ-II has been employed (see 1.7.1). This therefore introduces problems when considering the strength of these findings.

The inverse of acceptance is experiential avoidance, which has been found to engender a narrowing of one's behavioural repertoires, leading to a less values-consistent way of living and psychological distress (Bond et al., 2011). As experiential avoidance is considered to be any means of altering the form, frequency or intensity of private experiences, it includes a broad range of overt and covert behaviours including: thought suppression; distraction; daydreaming; alcohol and substance abuse; engaging in high risk sex; and overspending money (Zettle, 2007). Hayes et al. (1999) note that experiential avoidance can be adaptive in particular contexts (e.g. being able to avoid painful memories in work environments that require focus and concentration is important). However, when employed indiscriminately regardless of the context, experiential avoidance can become ineffectual or even harmful. Moreover, its intermittent effectiveness at reducing anxiety and other aversive internal experiences means that experiential avoidance becomes more resilient to extinction and thus pervades, even when it no longer works or causes harm. A clinical example of this issue can be seen in clients who utilise self-destructive methods (e.g. alcohol, drugs etc.) as a means to suppress aversive emotions such as grief. Grief cannot be avoided long-term as the stimuli which are used to avoid grief begin to participate in a verbal relation with it; in effect, distractions serve as reminders to the person of what they are distracting themselves from.

Despite this, the person may continue to use alcohol or drugs, even when these stimuli no longer permit experiential avoidance and may actually cause secondary problems (e.g. health-related problems, contact with the police, etc.)

The role of experiential avoidance as a mediating factor in treatment outcome has been supported across a number of clinical presentations including: social anxiety (Kashdan et al., 2014); chronic pain (Ruiz-Párraga & López-Martínez, 2015); post-traumatic stress disorder (PTSD; Boeschen, Koss, Figueredo & Coan, 2013); eating disorders (Fulton, Lavender, Tull, Klein, Muehlenkamp & Gratz, 2012); and depression (Cribb, Moulds & Carter, 2006). Two empirical reviews also support the role of experiential avoidance as a factor in the aetiology of a diversity of clinical presentations (see Chawla & Ostafin, 2007; Hayes et al., 2006).

1.4.2. Defusion - fusion. If fusion is the process by which covert private verbal events come to dominate overt behaviour over direct contingencies, defusion is the 'disentangling' of oneself from these events, and observing them for what they are; transient mental activity with no objective truth (Harris, 2007). ACT's defusion interventions aim to undermine the language processes which enable internal stimuli to carry the same aversive functions as the actual events they represent (Blackledge, 2007). More specifically, such interventions seek to change the functional contexts which promote the transformation of stimulus function and endow internal private events with the same potency as their externally located counterparts (Twohig, Masuda, Varra & Hayes, 2005). As an example of this, consider a person who experiences persistent negative self-evaluative thoughts such as 'I am ugly'. Fusion with such verbal content may lead to the individual avoiding situations where physical appearance is salient (e.g. social gatherings), which may then mean missing out on a valued aspect of life (e.g. the opportunity to form romantic relationships with others). In this example, fusion with the self-evaluative thought lends credence to its believability, and negates the chance to see this thought as a thought that is neither true nor false. A defusion exercise would aim to undermine the functional context which gives such a thought its emotional and personal meaning (e.g. by saying the words 'I am ugly' in a different accent, or saying it

rapidly and repetitively, imagining it has a colour, shape and texture etc.). The intended outcomes of this defusion exercise are: 1) a greater attention to the process of one's thinking, rather than the content; 2) an ability to question the believability of one's negative self-evaluate thinking; and 3) an increase in values-consistent behaviour, that is guided by direct contingencies rather than aversive internal stimuli (i.e. self-critical thoughts).

ACT notes that both acceptance and fusion with internal experiences will oscillate over the course of time (Flaxman et al., 2010). Indeed, despite one's willingness to embrace aversive private events, such events remain aversive and with greater fusion comes a concomitant decrease in willingness. Long learning histories, and a verbal 'community' which continues to support the literal use of language, mean that fusion processes cannot always be overridden (Hayes et al., 1999). Thus, aversive internal events will come to guide behaviour at some point. However, ACT suggests that this is not an issue per se, as long as there is a greater level of engagement in values-based behaviour, and less frequent experiential avoidance of internal aversive events overall.

Cognitive defusion has been implicated as a mediating factor in long-term outcomes for: anxiety disorders (Arch et al., 2012); depression (Zettle, Rains & Hayes, 2011); pain-related disability (Wicksell, Renöfalt, Olsson, Bond & Melin, 2008); substance abuse (Varra, Hayes, Roget, & Fisher, 2008); tinnitus (Hesser, Hayes, Westin & Andersson, 2008); professional burnout (Hayes Strosahl, Bunting & Twohig, 2004); and distress in response to hallucinations (Gaudiano, Herbert, & Hayes, 2010). Defusion strategies have also been found to produce a greater reduction in both the discomfort and the believability of negative thoughts compared to distraction and thought suppression tasks (Masuda, Hayes, Sackett & Twohig, 2004). An experimental test of a cognitive defusion exercise also illustrated that negative statements created less psychological discomfort, and participants were more willing to believe and engage with their content, when presented in a defused versus non-defused format (Healy, Barnes-Holmes, Barnes-Holmes, Keogh, Luciano & Wilson, 2008).

1.4.3. Self as context - self as content. The ability to adopt an observing perspective and notice one's internal experiences, rather than being engrossed within and self-defined by their content, is a key focus within ACT referred to as self as context (Hayes et al., 2004). ACT uses metaphors and experiential and mindfulness exercises to attempt to foster a distinction between the self who is noticing one's internal experiences and the experiences themselves. In this way, there becomes a psychological flexibility and freedom to live without being hampered or defined by limited descriptions of oneself (Bach & Moran, 2008). Indeed, without this observing perspective, the contents of private internal experiences can come to be self-defining characteristics through the relation framing processes of language (Strosahl, Hayes, Wilson & Gifford, 2004 pp. 44-45). This 'self as content' perspective (also referred to as 'conceptualised self') is considered the basis for the emergence of psychological distress in the ACT model.

Hayes et al. (1999) suggest that the emergence of the conceptualised self inevitably follows from a learning history of being taught to name, categorise, differentiate and evaluate ourselves, which accompanies the acquisition of language. A verbal community teaches and reinforces our ability to relationally frame our external and internal experiences with our self-definition of 'who we are', and generate narratives of our life histories (Hayes, 1995). However, where the self-stories held are pejorative and involve self-denigrating descriptions or causality-based assumptions (e.g. 'I'm a bad mother because I drink too much'), attachment to these stories can lead to distress in a number of ways. Firstly, there may be an increased overt avoidance of the contexts which promote negative self-defining stories, and potentially a loss of contact with aspects of life that are deeply valued. For example, a child may skip school if they define themselves as 'stupid' and perceive themselves as academically incompetent. This avoidance may also result in a loss of contact with things the child values as important in the immediate context (e.g. friendship groups), thus resulting in further distress.

There may also be attempts to experientially avoid these self-defining stories, particularly when we are fused with them, through self-destructive ways which then become the basis for further problems (e.g. self-harming as a means to avoid painful thoughts may result in an admission to a psychiatric inpatient unit, and thereafter impositions on one's freedom). Alternatively, distress may also emerge from attempts to maintain such negative self-stories despite their aversive connotations. RFT explains this motivation in terms of 'threats to the conceptualised self' being relationally equated with 'threats to me', which can then result in aversive emotions such as fear or anger (Törneke, 2010). The consequence of this is that other alternative and perhaps more positive narratives of the self are not sought, and may actually be avoided or distorted in order to fit with more historical and longer held narratives. In traditional CBT, this process of attempting to maintain consistency between our conceptualised self and external events by cognitive distortion is central to psychological distress, and is treated through cognitive restructuring to alter the content of negative-self stories (Clark, 1995). However, ACT focuses on attempting to weaken the attachment to such stories and develop a transcendent observing self who can notice such stories for what they are, and not be defined by them or act on them (McHugh & Stewart, 2012).

In comparison to the other processes within the hexaflex model, there appears to be relatively few studies within the literature that have sought to investigate the role of the self as context process in isolation. This may be in part due to the fact that, at present, there is only one specific measure of self as context (see Gird & Zettle, 2013) which has received little investigation of its psychometric properties. Theoretically however, there is support for the key role of self as context perspective taking skills when applied to clinical populations such as people with autism, who are thought to have difficulties with flexibly taking different perspectives (Rehfeldt, Dillen, Ziomek & Kowalchuk, 2007). Research also suggests that ACT interventions which incorporate self as context exercises may be more effective than those which do not (Williams, 2006). Levin, Hildebrandt, Lillis and Hayes' (2012) meta-analysis of ACT treatment component studies also found that self as context treatments (referred to as 'mindfulness' within the review) had a small but significant effective size on

targeted outcomes (*g* = .46), when compared to inactive conditions. However, interventions targeting other ACT processes were found to have a large effect size (*g* = .81, acceptance; *g* = .78, values plus mindfulness) and a medium effect size (*g* = .77, defusion; *g* = .64, present moment awareness) thus suggesting these processes may play a greater role in mediating psychological flexibility and distress. The findings also suggest that the effects of interventions targeting present moment awareness (see 1.4.4.) are significantly enhanced when also targeting values simultaneously; this reiterates the notion of ACT processes as being inextricably linked to each other (Hayes et al., 1999).

1.4.4. Present moment awareness - present moment absence. Contact with the present moment (or 'present moment awareness') is described as the shifting of attention to what is happening in the here-and-now, thereby making contact with both internally and externally located events or stimuli (Fletcher & Hayes, 2005). Present moment awareness also involves an 'attitudinal' component described as an openness and acceptance of experiences without judgement by “experiencing events fully and without defence, as they are” (Hayes, 1994, p. 30). Thus both awareness and acceptance in conjunction are considered to form this construct (Cardaciotto, Herbert, Forman, Moitra & Farrow, 2008). There also appears to be a conceptual overlap between ACT's present moment awareness and the perhaps more ubiquitously used construct of mindfulness, which has been defined as:

A process of regulating attention in order to bring a quality of non-elaborative awareness to current experience and a quality of relating to one’s experience within an orientation of curiosity, experiential openness, and acceptance (Bishop et al., 2004, p. 234).

This definition represents Bishop et al.'s (2004) attempt to provide the first formal scientific operation of Kabat-Zinn's more informal introduction of mindfulness into the psychological treatment vernacular as "paying attention in a particular way, on purpose, in the present moment, and non-judgmentally" (Kabat-Zinn, 1994, p. 4). However, the concept of mindfulness has been operationalised in several different ways with some defining mindfulness as a

repertoire of skills and techniques (reference), whilst others emphasize that mindfulness represents a transient mental state that can be harnessed within therapy for curative effects (for an overview see Brown, Ryan & Creswell, 2007).

Mindfulness strategies are incorporated into a number of psychological treatment packages in different forms or functioning including: mindfulness-based stress reduction (MBSR; Kabat-Zinn, 1982; 1998); MBCT (Segal et al., 2002); and DBT (Linehan, 1993). Generally, across each treatment package, mindfulness strategies are used to promote greater contact with and acceptance of aversive internal experiences, such that they do not influence overt behaviour in the same way. In ACT, mindfulness strategies are specifically used to cultivate present moment awareness, but with the specific purpose that this better enables behavioural engagement with one's values[11]. Put differently, mindfulness within ACT is a means to an end, with the end being behaviourally oriented and related to pursuing those aspects of life which have most meaning to a person. ACT does not use mindfulness interventions in order to directly reduce symptoms of psychological distress (as may be the case in MBCT and DBT), but this may be a secondary consequence to enhancing a more values-consistent way of living. Other mindfulness practices not founded within a psychological framework but within Buddhist philosophy (e.g. Vipassana and Zen meditations) similarly focus on promoting a state of contentment, rather than on directly reducing distress symptoms (Gilpin, 2009). However, mindfulness meditations diverge from the focus on behavioural modification within psychological treatments such as ACT, DBT, MBCT and MBSR, and instead focus on modifying a person's 'direct perception of pure experiences' (Rapgay & Bystrisky, 2009)[12]. Mindfulness meditations also teach meditation to provide the mental stability and clarity considered foundational within these traditions to achieving a mindful state. Treatments such as ACT however, do

[11] 'Self as process' is an alternative way of describing present moment awareness in relation to the 'self' within the ACT model. This involves an ongoing experiencing and describing of one's internal experiences in the 'here and now' (e.g. "I'm noticing I'm having the thought/feeling/sensation..."). The aim of doing this that one becomes less fused with and therefore less likely to be guided by such experiences (as is the case in self as content; McHugh & Stewart, 2012).

[12] However, it may be suggested that perception is also a (covert) behaviour itself. Thus, despite their purported different aims, mindfulness interventions in psychological treatments and mindfulness meditations both modify behaviour.

not incorporate meditation into their intervention packages, and suggest that meditation is unnecessary to practice mindfulness (Hayes & Shenk, 2004).

Flaxman et al. (2010) suggest that an absence of present moment awareness is intrinsic to the processes of evaluating, categorising, comparing and planning that occurs through verbal behaviour. Such processes not only occur after becoming aware of the event (and are therefore not 'present moment'), but they also involve focussing on events in the past (e.g. ruminating, reminiscing etc.) and the future (e.g. planning, worrying etc.) rather than the present. When such events contain information about aversive experiences, the ability to make contact with them in the absence of direct stimuli not only broadens the interface with painful experiences, but it negates an ability to respond to events in the present. From a behavioural analytic perspective, inattention to the present moment means that behaviour is less likely to be shaped by direct contingencies, as immediate reinforcing and punishing consequences are not noticed. Instead, inattention means that behaviours are more likely to be rule governed[13] and shaped by verbal stimuli (e.g. thoughts). In this way certain behaviours are not registered as either effective or ineffective in a given context, which may set up a cycle of repeatedly making the same mistakes (Luoma, Hayes & Walser, 2007). A lack of a present moment awareness therefore means that behaviour is more likely to be under the control of experiential avoidance and fusion with negative self-evaluate thoughts, rather than by a person's direct experiences (Hayes, Wilson, Gifford, Folette & Strosahl, 1996).

It should be noted however, that increased awareness and contact with private experiences, particularly those that are aversive, is not adaptive per se. Evidence suggests increased awareness in the absence of an attitude of acceptance is likely to exacerbate experiential avoidance (for a review see Hayes et al., 2006). Therefore, awareness in conjunction with acceptance is

[13]Behaviour that is under the control of verbal descriptions of contingencies rather than the direct contingencies themselves is referred to as rule governed (Törneke, 2010). RGB means that rules (e.g. 'look both ways when you cross the road, otherwise you will be hit by a car and die') can influence behaviour (i.e. looking both ways before crossing the road) without ever had any prior contact with the contingencies the rule describes (i.e. being hit by a car due to not looking both ways). Whilst this has positive consequences for our capacity to learn efficiently without the need for trial and error learning, the dominance of RGB over direct contingencies of reinforcement is central to ACT's perspective of the development of psychological distress. For a more detailed discussion on this see Törneke, Luciano and Valdivia-Salas (2008).

arguably crucial for mindfulness interventions in ACT and other therapies to have a positive rather than aversive therapeutic effect (Cardaciotto et al., 2008). This contrasts with more traditional behavioural interventions based on exposure and habituation principles, as contact with aversive external events is considered important for sensitisation to aversive stimuli.

The absence of present moment awareness (or 'mindfulness') has been implicated as a mediating factor in the development and maintenance of dissociation and PTSD symptoms following traumatic events (Boughner, Thornley, Kharlas & Frewen, 2016). Broadly, dissociation has been defined by the *Diagnostic and Statistical Manual of Mental Disorders* as a disruption of the usually integrated functions of consciousness, memory, identity or perception of the environment (5th ed.; *DSM–5*; American Psychiatric Association [APA], 2013).

Dissociative states are evident in a number of psychological disorders such as PTSD, major depressive disorder, borderline personality disorder and schizophrenia. However, a lack of conceptual clarity and agreement about what constitutes 'dissociation' has meant that the term has been used to cover a wide range of behaviours and experiences, with varying degrees of psychological distress developing as a consequence of the dissociative process (Kennedy, Kennerley & Pearson, 2013). If conceptualised on a continuum, dissociative experiences at the 'lower' (i.e. non-clinical) end can include ubiquitous and relatively harmless experiences such as daydreaming (e.g. being on 'autopilot' whilst driving and not remember how one got to a particular destination). At the more clinical and severe end of this continuum, dissociative experiences can include amnesia, depersonalisation and emotional 'numbing', intrusive thoughts and hallucinations, and the development of multiple personalities (dissociative identity disorder [DID]) (Dell & O'Neil, 2010).

Despite these differences research suggests that these more clinically distressing dissociative states typically develop following traumatic incidents (van der Kolk, McFarlane & Wiesaeth, 1996) in order to manage overwhelming emotions and fear connecting to these events historically.

Evidence suggests that dissociation is a key mediator of psychological distress following trauma events in a diversity of different clinical groups including: victims of intimate partner violence (DeCou, Lynch, Cole & Kaplan, 2016); voice-hearers (Pilton, Varese, Berry, & Bucci, 2015) and women with non-suicidal self-injury (Franzke, Wabnitz, & Catani, 2015). Together with the aforementioned evidence regarding the mediating role of present moment awareness in the development of dissociation, this evidence supports the rationale for interventions that address dissociation by promoting present moment awareness. Indeed, whilst numerous theories have since provided different aetiological accounts of dissociation and led to different treatments (for an overview see Putnam, 1997), contemporary treatments of dissociative states such as trauma-focussed CBT (Ehlers & Clark, 2000) and eye movement desensitisation and reprocessing (EMDR; Shapiro, 1989a) share a focus on helping people to re-orientate themselves to the 'here and now' rather than disengage. Acceptance-oriented therapies, such as ACT, also do this, but arguably in a more explicit manner because of the model's conceptualisation of distress emanating from a deficit in present moment awareness.

ACT views dissociation as a form of experiential avoidance, whereby aversive internal experiences (thoughts, feelings, memories) acquire their aversive functions through fusion (i.e. a lack of distance between the content of these experiences and oneself), leading to both overt and covert behavioural attempts to disengage (i.e. dissociate) from such experiences (Kennedy, Kennerley & Pearson, 2013). ACT suggests that the result of this dissociative disengagement is a lack of present moment awareness to the 'here and now', and an inability to effectively respond to present contingencies in a way that leads to a rich and meaningful life. ACT treatment for dissociative states therefore focuses in particular on cultivating a person's present moment awareness to: (1) counter unconscious disengagement; and (2) promote a conscious engagement with current experiences (internally and externally), whether pleasant or aversive, in an open, curious and non-judgemental way (Harris, 2009).

Whilst no specific ACT process measure of present moment awareness exists, ACT clinicians (such as Harris, 2009) promote the use of mindfulness measures to examine changes in this present moment awareness following ACT treatment for dissociation. Whilst advantageous, such process-specific measures do not additionally measure other important ACT processes which are also focussed on as a part of a broader treatment package for dissociative states; this again provides a rationale for the need of a general measure of ACT processes in clinical contexts other than the AAQ-II.

1.4.5. Values. Values is a central principle within ACT and aligns most closely to the therapy's behavioural theoretical origins and clinical orientation. Living a valued life is also the crux upon which ACT therapists motivate people to increase their contact with and acceptance of aversive experiences (i.e. informing clients that doing this will mean living a life that has personal meaning and currency) (Wilson & Murrell, 2004).

Hayes et al. (2012) describe values as "chosen, verbally constructed, consequences of patterns of activity, for which the predominant reinforcer becomes intrinsic to the behavioral pattern itself." (p. 40). The emphasis here is the notion of chosen behaviour, as behaviour that is guided by social compliance and/or avoidance of aversive internal and external events is under aversive control, and is therefore likely to perpetuate distress (Sidman, 2001). There is also an emphasis on the fact that what reinforces values-directed behaviour is engaging in the behaviour itself (i.e. process), rather than trying to achieve some distant goal (i.e. outcome) (Wilson & DuFrene, 2009). For example, if an individual's goal is to be a 'good parent', an ACT therapist would discuss what behaviours would likely constitute *being* a 'good parent' from that person's perspective. The purpose of this emphasis on process over outcome is related in part to the fact that outcomes are beyond the control of both the individual and the ACT therapist working with the individual to influence their behaviour. Conversely, the process of trying to behave in ways that will help achieve one's goals is more reliably under the control of the client, and therefore is more open to reinforcement by the ACT therapist (Flaxman et al., 2010).

ACT first examines a person's core values in a variety of domains (e.g. work, relationships, leisure etc.), and from this point begins to establish behaviours which correlate with these values, as well as the goal of such values (Harris, 2009). In the course of values work, discussions are held about what factors may inhibit values-consistent behaviour and how to overcome this. Where such factors are psychological in origin (e.g. experiential avoidance and fusion with aversive events), metaphors and experiential and mindfulness interventions are used to undermine the verbal processes which create these issues.

Working to enhance values consistent behaviour (or 'valued living') can be distinguished from the basic behavioural notion of reinforcing desirable and adaptive behaviours by the nature and location of reinforcers. Values are purported to be behaviours related to the subjective meaning and personal purposes of the individual, whereas behaviours unconnected to values are considered to relate more to predominant cultural rules (Flaxman et al., 2010). ACT argues that behaviour divorced from personal values and under the control of prescribed cultural decrees is more likely to be under aversive control (e.g. individuals behave in certain ways to avoid ridicule, disappointment and other forms of social punishment; in relation to this, see Torneke et al.'s, 2008 discussion on RGB and 'pliance' from an RFT perspective). However, it is recognised within ACT that personal values are likely to be influenced by prevailing cultural attitudes, and thus the aim is to ensure that such values also have a degree of personal meaning for the individual.

ACT promotes the idea that behaving in line with personal values is likely to lead to contact with aversive events and experiences at times. However, RFT suggests that due to being relationally framed in equivalence to one's verbally stated values, such behaviours may become reinforcing despite being aversive (Törneke, 2010). In more simple terms, a person may be able to behave in

ways that are immediately perceived as aversive, if such behaviour is considered as in-line with one's more deeply held personal values[14].

Evidence suggests that interventions which incorporate values work have positive therapeutic effects, particularly in the area of chronic pain where post-intervention changes in values-based action have been negatively correlated with changes in: pain intensity (r = -.30); depression (r = -.41); physical disability (r = -.39); and psychosocial disability (r = -.40) (Vowles & McCracken, 2008). Tolerance of pain was also found to be greater in participants receiving an ACT-values protocol than participants receiving a control-values protocol, with 70% of participants in the ACT protocol condition tolerated the maximum number of shocks compared to 10% of participants in the control protocol condition (Páez Blarrina, Luciano, Gutiérrez Martínez, Valdivia, Ortega, & Rodríguez Valverde, 2008). The combination of values interventions alongside acceptance interventions has also been found to enhance pain tolerance compared to acceptance interventions in isolation (e.g. Branstetter-Rost, Cushing, & Douleh, 2009). Values exercises can also help ameliorate stress responses (e.g. Creswell et al., 2005) and have been recommended to be incorporated into weight-control programs based on their effectiveness in this regard (see Lillis & Kendra, 2014 for a discussion). Values exercises have also been found to have positive effects in non-clinical areas including such as academic performance (e.g. Chase, Houmanfar, Hayes, Ward, Vilardaga, & Follette, 2013) and intimacy in romantic relationships (e.g. Lomore, Spencer, & Holmes, 2007).

1.4.6. Committed action - inaction and impulsivity. Hayes et al. (1999) describe committed action as "values-based action designed to create patterns of action that is itself values-based" (p. 95). Thus there is interdependence between values and committed action processes within the ACT model. Ostensibly however, values and committed action processes may be difficult to distinguish when considered from a behavioural perspective. Values are a verbal description of behaviours that are reinforcing for an individual based on their learning history (functionally values have been defined

[14] For example, reprimanding one's child for irresponsible behaviour may initially appear harsh if the child becomes upset and cries, but such behaviour may be reinforced by a concomitant perception of this behaviour as that of a 'good parent' by teaching the child values and morals.

as 'augmental verbal rules' or 'establishing/motivating operations'; see Plumb, Stewart, Dahl & Lundgren, 2009). Committed action is the behavioural enacting of such verbal descriptions. When viewed through a behavioural 'lense', both values and committed action are at a most fundamental level different forms of behaviour, which are both subject to contingencies of reinforcement. In other words, values behaviour and committed action behaviour are different topographically, but functionally may be related. Within the ACT literature, these processes are differentiated based on the idea that it is possible to have values without committed action. For example, a person can state that they wish to be a 'good father', but fail to act in ways consistent with this self-definition, such as attending parent's evenings, spending time with their child instead of working, etc. This person may continue to verbally state their 'desire' to be a 'good father', (as such statements may be reinforcing due to societal values associated with parenthood) without ever translating this 'desire' into tangible action; this process of translation is a core feature of ACT therapist's values-based interventions (see Harris, 2009).

Flaxman et al. (2010) help to tease apart and operationalise committed action as involving the following elements: 1) the enacting of one's values, as opposed to simply stating that one will behave in a certain way; 2) the moment by moment awareness of when behaviour has become values-inconsistent, and the subsequent re-direction of behaviour to re-align with one's values; 3) committing to act in a values-consistent way, even when doing so may result in contact with aversive experiences; 4) adopting an attitude of acceptance and taking the next opportunity to act in a values-consistent way (depending on what the situation affords) where attempts to behave in a values-consistent way fail; 5) publicly declaring one's commitment to one's values, such that help may be enlisted from others in this regard; and 6) planning how one will meet long-term values by setting intermediary goals, and describing how one will achieve them and overcome potential barriers.

Point 5. above shows that where values and committed action are the focus, ACT utilises a number of traditional behavioural change methods including goal setting, skill-development, and exposure. However, Hayes et al. (2012) suggest

that the focus of such methods when employed in ACT is different to traditional behavioural treatments, because there is an emphasis on enhancing commitment to values-consistent action in a context of openness and awareness; that is, there is a fostering of a curiosity and non-judgemental stance to oneself if values are not committed to, such that one can learn about the potential barriers to values-consistent action in the future. Evidence also suggests that traditional behavioural interventions can be enhanced by incorporating ACT processes, such as acceptance (see Dahl, Wilson & Nilsson, 2004; Eifert & Heiffner, 2003; Levitt, Brown, Osillo & Barlow, 2004; Wicksell, Ahlqvist, Bring, Melin & Olsson, 2008).

In the absence of committed action, two overarching consequences are possible: either a narrowing of behavioural repertoires (i.e. inaction, passivity and withdrawal); or an excessive of unhelpful behaviours (e.g. impulsive behaviours such as drinking, taking drugs, self-harming etc.). People engaged in these forms of behaviour are described within the hexaflex model as lacking psychological flexibility to act sensitively to direct contingencies of reinforcement. Instead, they are more likely to be guided by indirect private stimuli such as thoughts, feelings, memories and sensations. Without this flexibility, ACT suggests that the ability to adapt and change one's behaviour to fit with the current circumstances is limited, and a less values-consistent way of living is more likely to result.

1.6. ACT's evidence base

Six meta-analyses (Hayes et al., 2006; Öst, 2008; 2014; Powers, Zum Vörde Sive Vörding & Emmelkamp, 2009; Ruiz, 2012; Tjak, Davis, Morina, Powers, Smits & Emmelkamp, 2015) have investigated ACT's efficacy across a range of clinical presentations, and there remains a lack of consensus as to the strength of ACT's evidence-base.

Hayes et al. (2006) reported on 21 RCTs of ACT and found an average medium effect size at post-treatment (d = .66) and at follow-up on average 19.2 weeks later (d = .65). When compared to active treatments, Hayes et al. found ACT to have a small effect size post-treatment (d = .48), but this increased to a medium effect size at follow-up (d = .62). In comparison to waiting lists, treatment as

usual, or placebo treatments, ACT interventions demonstrated a large effect size post-treatment (*d* = .99) and at follow-up (*d* =.71).

Öst's (2008) meta-analysis compared 13 RCTs in ACT with traditional CBT and concluded that due the lack of methodological rigour in the ACT studies, ACT could not be classified as an 'empirically supported treatment' [15]. In a rejoinder, Gaudiano (2009a) argued that Öst's meta-analytic methodology was cofounded by not comparing like-for-like samples in the ACT and CBT studies (e.g. comparing ACT's studies into treatment-resistant populations with CBT's trials of emotional disorders). Gaudiano (2009b) also later suggested that Öst's methodological rating system would unfairly downgrade any new and emerging psychotherapies, such as ACT, on the basis of the amount of grant funds available to support the trial, the particular stage of treatment development at the time the study was conducted, and the specific population being studied.

In a later meta-analysis and systematic literature review, Öst (2014) suggested that there has been no significant improvement in the methodology of ACT studies compared to his 2008 meta-analysis, and there had actually been a deterioration in the effect size of studies relating to ACT (from *d* = .68 to .42). Öst also found a small and non-significant effect size when comparing ACT to a number of cognitive or behavioural forms of treatment, and concluded that ACT is not a well established for any disorder. Conversely, Smout, Hayes, Atkins, Klausen and Duguid (2012) suggest that since Öst's (2008) initial meta-analysis, ACT studies have improved their methodological rigour (e.g. by monitoring treatment adherence and fidelity), and findings generally support ACT's efficacy across a number of presentations including: chronic pain; obsessive–compulsive disorder; and a number of anxiety disorders (including panic disorder, social phobia, and generalised anxiety disorder).

Powers et al.'s (2009) meta-analysis of 18 RCTs found a clear overall advantage of ACT compared to control conditions (i.e. non-active and treatment as usual groups; *g* = .42), but found no evidence to support ACT as more beneficial compared to well established therapies such as traditional CBT

[15] For a review of the 'empirically supported treatment' concept, and the controversies surrounding it, see Chambless and Ollendick (2001).

(*g* = .18; *p* = .13). However, a re-analysis of Powers et al.'s data by Levin and Hayes (2009) concluded that ACT was better than these established treatments (*g* = .27; *p* = .03).

Finally, Tjak et al. (2015) reviewed 39 randomised control trials (RCT) examining ACT's efficacy in relation to mental health disorders and somatic health problems. Whilst ACT was not considered to be statistically superior to CBT (*p* = .14), it was more effective than waiting list (*g* = .82) treatment as usual (*g* = .64) and placebo controls (*g* = .51), and was as effective as CBT for the treatment of depression, anxiety disorders, addiction and somatic health problems.

It is worth considering that the evidence for treatment efficacy within these studies is related to treatment outcomes, and does not provide evidence attesting to the processes through which such treatment packages have been effective. Levin and Hayes (2009) argue that in order to understand what treatments are effective and why, empirical research must broaden its standards and method of investigation by including meditational analyses in outcome studies, as well as 'gold standard' RCTs and EST criteria. Ruiz (2012) examined 16 studies which compared ACT and CBT interventions across different outcomes, and found that mean effect sizes on primary outcomes significantly favoured ACT over CBT (*g* = .40), except on depression and anxiety outcome measures where there was no statistically significant differences between therapies[16] (*p* = .12-.14). Importantly, Ruiz also found that out of the nine reviewed studies which conducted mediation analyses, six studies supported ACT as working through its purported processes of change (i.e. decrease in experiential avoidance and cognitive fusion processes), whereas four studies failed to find CBT working through its hypothesized processes of change (i.e. change in dysfunctional attitudes and reduction in the frequency of negative automatic thoughts). This finding reiterates Longmore and Worrell's (2007) suggestion that the mechanisms of change through which CBT works by remain unclear. However, conclusions about evidence supporting ACT's theoretical processes over CBT's in these studies should be considered

[16] Ruiz (2012) suggests however that the baseline levels of depression were lower in the CBT samples, thus suggesting ACT is marginally superior for the treatment of depression.

in light of the limitations of the measures used to examine changes in ACT processes (see 1.7.1.).

Despite mixed findings for its efficacy, ACT has been recognised as an empirically supported treatment by the American Psychological Association (APA, 2006) and has been moderately supported as applicable for people experiencing depression, and strongly supported for people experiencing chronic pain. ACT is also listed as an evidence-based practice on the Substance Abuse and Mental Health Services Administration's (SAMHSA, 2010) national registry of evidence-based programs and practices. However, traditional CBT's hegemony as the recommended form of treatment for a range of mental health issues remains in the United Kingdom.

1.7. Rationale for the current study

1.7.1. Limitations of the current general measure of ACT processes.

Wolgast (2014) argues that the AAQ-II fails to operationalise items in a functional contextual consistent manner. Specifically, the items address how possessing psychological (in)flexibility to varying degrees (in a 'trait' like manner) impacts on *how a person responds* to internal events which results in psychopathology. Wolgast asserts that this problem echoes a core criticism of the AAQ, that the measure confounds process or trait with outcome. This makes it difficult to delineate whether "a specific response is grounded in levels of psychological inflexibility/experiential avoidance or, for example, in levels of experienced aversive emotions, memories, and worries." (p. 833). As a result, responses to items at both the level of psychological inflexibility or at the level of experienced aversive internal experiences are likely to receive the same score within the AAQ-II, despite the fact that only the first aspect of this is being targeted by the measure.

As an example of this issue, item six of the AAQ-II states 'it seems like most people are handling their lives better than I am'. Whilst ostensibly this item appears to try and tap into the ACT process of fusion, implicitly this actually taps into a distress outcome (i.e. the result of comparing oneself to other people)

rather than examining the (fusion) process by which such comparisons occur. There is therefore a risk in the circularity of such items within the AAQ-II.

Wolgast (2014) also questions the discriminant validity of the AAQ-II based on research findings that items from the measure both load onto the same factor as items designed to measure general distress, whilst also failing to load onto the same factor as items developed to measure acceptance/non-acceptance/psychological flexibility. Gámez et al. (2011) argue that this issue derives from the overlapping content and focus on 'distress leading to dysfunction' within both the AAQ-II's items and items within distress measures. Therefore, when attempts are made to evidence the AAQ-II's discriminant validity through correlations with theoretically distinct outcome measures such as distress measures, the overlapping content of each measure's items confounds and thus 'contaminates' this evidence.

In addition to the aforementioned criticisms, I would also add that the problem of the AAQ-II's poor discriminant validity stems from fundamental issues with the face and content validity of many of the AAQ-II's items, which resultantly conflate distress with process. Indeed, items two ('I'm afraid of my feelings') three ('I worry about not being able to control my worries and feelings') and seven ('worries get in the way of my success') appear to target distress and meta-distress constructs (i.e. 'worry', 'worry about worry', and/or 'fear of feelings'), rather than ACT-related processes. Additionally, there is also a preponderance of items which focus on the ACT processes of acceptance and fusion, but no items appear to examine other processes in the hexaflex model such as present moment awareness, values, or committed action.

Whilst others such as Gámez et al. (2011; 2014) have attempted to address the limitations of the AAQ-II, the measures they have developed such as the Multidimensional Experiential Avoidance Questionnaire (MEAQ), and its briefer version the Brief Experiential Avoidance Questionnaire (BEAQ), appear to measure constructs that are considered related but ultimately distinct concepts within the ACT literature. 'Experiential avoidance' was the overarching term originally used to describe the ACT model. Experiential avoidance is defined as attempts to alter the form, frequency, or situational sensitivity of negative private

events (e.g. thoughts, feelings etc.), even when doing so leads to behavioural difficulties. Acceptance was therefore considered as the inverse of experiential avoidance and instead involved a person willingly experiencing aversive internal events in the pursuit of values and goals. However, as Bond et al. (2011) note, as ACT has developed over time there has been a more focus on other processes within the model. There has also been recognition that experiential avoidance may be actually be helpful where there is limited opportunity to pursue one's values and goals in a given context. Thus experiential avoidance and acceptance have gradually acquired a narrower meaning that no longer connotes with ACT's central ethos. Instead, psychological flexibility is the term now used to emphasize that ACT aims to promote a more flexible fit between the person's context, private experiences and valued actions. Gámez et al. developed the MEAQ, and then the BEAQ, to overcome the AAQ-II's issues, but even they suggest that their concept of experiential avoidance is not the same as 'psychological flexibility' espoused by Bond et al. Thus, there is arguably a need for a more theoretically consistent and general measure of ACT processes and psychological flexibility.

1.7.2. Six versus three core ACT processes. In the initial development of the ACT model, Hayes et al. (1999) suggested that six interrelated processes form the basis of psychological flexibility. However, the authors go on further to suggest that each process is theoretically more deeply interlinked with another particular process, and thus processes can be paired to form three 'response styles': 1) acceptance-defusion form 'open' responses; 2) present moment awareness-self as context form 'centred' responses; and 3) values-committed action form 'engaged' responses. Suggestion that ACT's six processes may be collapsed or combined into fewer processes was later again reiterated by Hayes et al. (2004), who categorised the ACT model into two superordinate components: *mindfulness and acceptance processes* (present moment awareness; self as context; acceptance and cognitive fusion); and *commitment and behaviour change processes* (values and committed action). Hayes et al. (2011) also argue that ACT's six processes could be clustered into three groups of dyadic processes (see Figure 2 below): 1) 'openness to experience and

detachment from literality' (acceptance and defusion); 2) 'self-awareness and perspective taking' (present moment awareness and self as context); and 3) 'motivation and activation' (values and committed action).

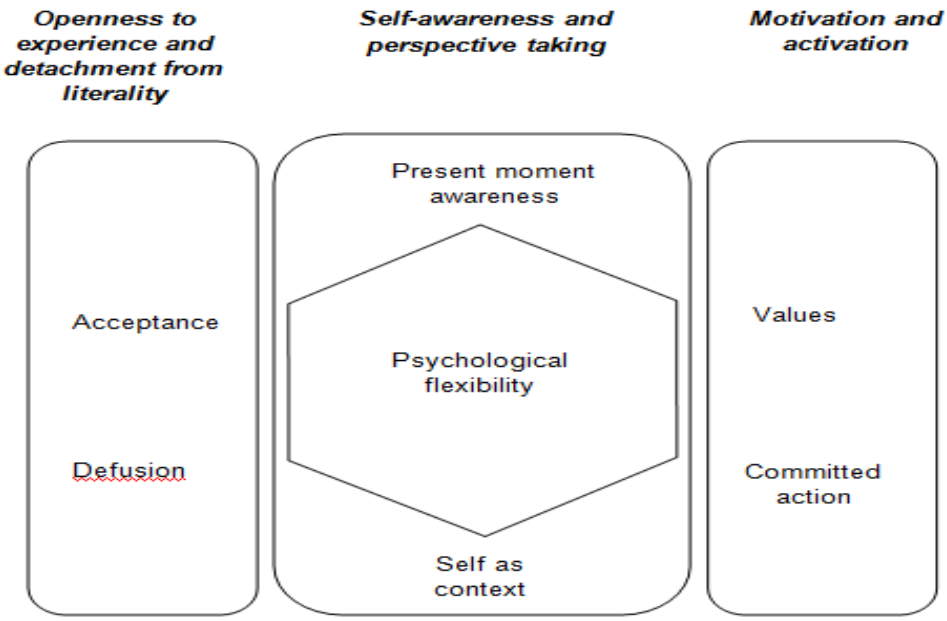


Figure 2. Hayes et al.'s (2011) clustering of ACT processes into three dyadic groups that make up psychological flexibility.

2. Extended methods

2.1. Phase one: Measurement development

2.1.1. Generation of initial item pool: Selecting measures. DeVellis (1991) argues that at this stage it is important that the breadth of the constructs to be measured are fully captured through having a large number of items. Thus, it is better to err on the side of caution and be over-inclusive in terms of the number of items rather than prohibitive at this stage. However, as our study involved ACT experts devoting a considerable amount of their personal time, I were conscious of minimising the risk of participant fatigue and attrition by not being too over-inclusive. In total, 106 items were selected for the initial item pool (see Appendix B). Before considering the procedure by which these items were

selected from existing measures, I will detail the three main ways in which the measures themselves were identified:

1) Ciarrochi and Bilich (2006) have compiled a directory of ACT process-specific measures, some of which are directly derived from the ACT model, and others which measure psychopathology constructs that may overlap theoretically with ACT processes[17]. However, in order to ensure fidelity to ACT theory, only those measures which are directly derived from the ACT model were selected from Ciarrochi and Bilich's list to extract items from.

2) A literature search was performed through Google Scholar from 2006 onwards to identify ACT measures developed since Ciarrochi and Bilich's (2006) ACT process measurement directory was compiled. A similar search was also performed in the Journal of Contextual and Behavioural Science's (JCBS) database. A search for 'acceptance and commitment therapy measure' was performed, and only those studies which were published in English were included.

3) ACT experts were asked to suggest ACT process-specific measures which were not found by the above methods. This was done by posting a request on a discussion group for 'ACT professionals' hosted on the Association for Contextual Behavioural Science's (ACBS) website. A list of the measures already obtained was included within this post in order to prevent duplication and repetition.

Based on the above steps, 17 ACT process measures were initially identified. Each measure was then reviewed to check whether its psychometric properties were supported by a number of studies, and whether it appeared to measure the ACT processes of interest (i.e. its face validity). Six measures failed to meet this criteria, and thus 11 measures were selected from which items could be extracted from (see journal paper Table 1).

2.1.2. Generation of initial item pool: Item selection procedure. Items were chosen principally in terms of their factor loadings, as this provides a

[17] For example, Ciarrochi and Bilich (2006) suggest that measures of narcissism (e.g. Narcissistic Personality Inventory [NPI]; Raskin & Hall, 1979) involve examining the believability of a person's inflated self-evaluations, and thus may be analogous to the ACT process of defusion/fusion.

metric of the relationship between surface variables (e.g. responses to questions on a measure), and the underlying latent variables (or factors) they are purported to measure (Comrey & Lee, 2013). A higher factor loading indicates a stronger relationship between an item and an underlying latent variable, and thus provides indirect evidence of an item's construct validity. Where there were several items with the same factor loadings within a measure, all such items were included in the initial item pool.

The factor loadings for each item were based upon data relating to non-clinical samples. Whilst many of the measures also established item factor loading data in relation to clinical samples, I did not select items based on this information as I intended to initially validate using a non-clinical sample and introduce a clinical sample at a later date. The graduation from a non-clinical to a clinical validation sample appears to be a commonly used procedure in the validation of new psychometrics, and indeed many of the existing ACT process measures (including the AAQ-II) followed this progression.

Selecting items to be included within a measure simply based on item factor loadings may be problematic however, as loadings give an indication of how items statistically 'hang together', and tell us little regarding their conceptual accuracy (DeVellis, 1991). Thus, whilst the majority of items were selected using the above statistical criteria, there were a number of exceptions to this:

1) Some items were selected if they had good face and content validity, even if they had lower factor loadings than other items. Given our earlier criticisms of the face and content validity of items within the AAQ-II, I did not wish to make a similar mistake when developing our measure.

2) All items from the BEAQ were included in the initial item pool. This was done as I had concerns that there were too few items to represent the full breadth of ACT's acceptance process. Indeed only four empirically supported of measures of acceptance were identified within the literature, and the paucity of such measures is perhaps surprising given the centrality of this process to the ACT model. Although I initially considered examining whether items from measures theoretically related to ACT's acceptance process could be included in the initial

item pool, I ultimately decided to avoid this in order to ensure fidelity to ACT's theoretical model.

Importantly, additional acceptance items were not selected from the MEAQ, as the BEAQ is a shortened version of the MEAQ that was developed based on the removal of items with the lowest factor loadings. Additional acceptance items were also not chosen from the AAQ-II given the issues with the measure's discriminant validity, and the poor content and face validity of the measure's items.

3) Items selected to represent the ACT dyadic process of 'self-awareness and perspective taking' (combining self as context and present moment awareness) in the initial item pool were all taken from the Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003). Mindfulness is a construct that has been attested to overlap in particular with ACT's present moment awareness process (see 1.4.4.). At present there are no specific ACT measures of present moment awareness. Thus it was decided that items representing this construct should be taken from mindfulness measures given their conceptual overlap. A recent systematic literature review identified the MAAS as the most psychometrically robust measure of mindfulness (Park, Reilly-Spong & Gross, 2013). Given the plethora of mindfulness measures within the literature, I considered choosing items only from the MAAS as helpful in our attempts to prevent participant burden and attrition associated with having to rate too many items.

Additionally, only one self as context measure (see Gird & Zettle, 2013) currently exists within the literature. This measure has not undergone an extensive validation process and as a result, I initially chose not include items from this measure in the initial item pool. However, some items from the measure were later included for rating during the Delphi consensus phase of the study, after the ACT expert panel suggested that the self as context aspect of this construct was not being fully captured by the existing items.

4) In addition to the above process, I and my supervisors also generated 37 items (see Appendix B for a list of our own self-generated items). The decision to supply our own items came after dissatisfaction with the face and content

validity of some of the items selected based on the statistical threshold criteria alone. The supervisors of this research are clinical and research psychologists with experience of writing and publishing papers on ACT, RFT and other areas theoretically related to functional contextualism. The items generated were therefore underpinned by a good theoretical and clinical understanding of the ACT model.

DeVellis (1991) stresses the importance of clarity and succinctness in relation to the wording of newly created items. However, Bond et al. (2011) note that ACT processes are "subtle construct[s] that can be difficult to convey in short statements that are understandable to people uninitiated in ACT..." (p. 10). Although I agree with Bond et al.'s statement, I also recognise the need to create items that are not unnecessarily wordy. In order to address this issue I sought to analyse and ensure the readability and comprehensibility of items in the final measure (see 2.1.6.).

I included a mixture of positively and negatively valenced items within the initial item pool. The decision to include of items with a mixed valence was based on our earlier criticisms of the AAQ-II, which only included negatively valenced items. I wanted to develop a measure that has items which when both endorsed and not endorsed suggests a high level existence of the latent variable (i.e. the particular ACT process). The inclusion of items of opposite valence can also help to prevent response acquiescence (i.e. the tendency of responders to provide affirmative answers to items irrespective of the item's content; Messick, 1991). However, research suggests that the existence of items of opposing valence in a measure can be confusing for respondents to complete, and can also impact on the measure's underlying factor structure (Credé et al., 2009). I intended to test out this former issue by asking participants to rate the face and content validity of items, meaning that those items which were considered poor or confusing would be dropped from the final measure. Thus, whilst I intended to have a composition of items with a mixed valence, ultimately our participant's ratings of items would have the decisive bearing on this. I also intended to test this hypothesis out by reviewing the stability of our measures factor structure.

2.1.3. Review of initial item pool by experts. Content validity is defined as the extent to which a measure or item represents the theoretical construct it is intended to reflect (Cronbach & Meehl, 1955). DeVellis suggests that the content validity of a measure can be maximised by involving experts in three key ways: 1) asking experts to rate how relevant they believe each item is to what it is supposed to measure; 2) asking experts to highlight aspects of the constructs of interest which are not being measured by the existing items; and 3) asking experts to evaluate each item's clarity and comprehensibility. I sought to involve ACT experts in all of these ways by using a Delphi-consensus methodology to develop our measure.

2.1.4. The Delphi method: A review of a consensus building approach. In the Delphi-consensus approach, data is collected from a 'panel' of experts through several rounds of rating items in a questionnaire that addresses a particular issue. The expert panel's responses are anonymised in order to minimise the effects of social influence and social conformity which have been found to affect decision-making in group contexts (see 4.1.). Responses from experts at each round determines which items are retained and which items are excluded in subsequent rounds based on a priori defined rating thresholds.

Traditionally, a statement on the positions of the experts is given to each panel members following the completion of each round. This 'controlled feedback' aims to promote a consensus of opinion amongst panel members by allowing them to see their fellow panel member's feedback, and then consider whether to change their ratings or not based upon this (Linstone & Turoff, 1975). There have been a number of adaptations to this process however, allowing for greater flexibility in light of the particular aims of researchers. For example, the provision of summary statements to panel members is not always used with the intention of eschewing a change in their opinions. Instead, such statements are used principally as a means to communicate decisions with panel members, which some suggest is crucial to ensuring their continuing motivation to engage with the Delphi process (Bosun & Modrak, 2014).

Of particular relevance to the current study, the Delphi method has guided the development and selection of measurement tools in mental healthcare settings:

this includes selecting the best self-report measures of mental health and wellbeing in child and adolescent mental health services (CAMHS; Deighton, Crudace, Fonagy, Brown, Patalay & Wolpert, 2014); developing quality control measures for child mental health service provision (Zima et al., 2013); and developing a measure of recovery from addiction (Neale et al., 2014).

2.1.5. Selecting ACT experts from the ACBS website. There were around 7000 members subscribed to the ACBS website at the time of phase one of our study. Many of these members did not fit our inclusion criteria for expert status. I therefore sought to refine our search criteria in the membership directory in order to retrieve suitable participants. This involved selecting 'counsellor/therapist/clinician' and 'researcher' under the 'profession' criteria. I also specified 'English' under the 'language spoken' criteria. 475 possible participants were retrieved based on this search strategy. From these 475 participants, I initially targeted those individuals who are well known in the ACT community as expert clinicians and/or researchers, as well as those who had already developed some of the measures from which items have been taken from for rating in our current study. The remaining participants were targeted for recruitment if their personal profile clearly indicated an expert status consistent with our inclusion criteria.

Emails were sent to the 50 prospective participants requesting their participation in the study (see Appendix D). An overview of the purpose and nature of the study was included within the email, as well as a link to the website hosting the survey. The initial page of the survey also provided more information about the study. Participants were required to indicate their consent by providing their name and an email address which they could be contacted through directly. After this, participants were then able to proceed to the next page of the survey and begin rating items.

2.1.6. Selecting experts for Delphi studies. The selection and composition of the expert panel in a Delphi study is considered crucial in generating high quality data (Taylor-Powell, 2002). However, there are currently no set criteria to inform how and how many experts should be selected and recruited within a Delphi-consensus study (Hsu & Sanford, 2007). Detailed

knowledge of the subject matter is considered essential, as well as the flexibility to adapt and change perspectives on an issue in the service of attaining consensus (Oh, 1974). Moreover, enough panel members to capture the nuanced differences of view of a subject matter are important, with a typical Delphi study having around 15 to 20 respondents (Ludwig, 1997). Some suggest however that there is no evidence indicating a correlation between the number of participants included within a Delphi study and the reliability and validity of the consensus building process (Murphy et al., 1998). Instead, the determinants of sample sizes in Delphi studies should be based on several factors, including: 1) the nature and scope of the problem; 2) the cost and time devoted to the rating process; 3) possible participant fatigue and burden relating to the number of items to be rated; and 4) total length of time to be committed to this process (Hasson, Keeney & McKenna, 2000). Ten ACT experts completed all rounds of rating during the Delphi phase of the study. All members identified themselves as clinical psychologists who had been practicing ACT for at least five years (observed range of 5-18 years) and had published ACT-related articles in peer reviewed journals. Of the ten members: three were certified ACT trainers and supervisors, one of whom had co-authored a published book-chapter on ACT processes; four identified themselves as professors/assistant professors; one was an academic director of a clinical doctoral programme within the United Kingdom; and two had been involved in developing (two of the) ACT process measures from which I extracted items used in this study (measures representing defusion/fusion and acceptance/avoidance processes).

2.1.7. Attrition of participants in the Delphi phase. Eight additional participants provided their consent to take part in the study in round one, but either failed to complete the survey, or they did not begin the survey at all. Participants were removed from successive rounds of ratings if they failed to complete the survey in the allocated two week schedule. This was done due to the time-limited nature of collecting data for our research. Email reminders were sent to those participants who had not completed the survey one week before the survey was due to close. A one week period was held between each round of rating to allow time for data analysis. The only exception to this was between rounds two and three when there was a two week interval, during which I and

my supervisors were creating additional items for rating based on the expert panel's feedback.

2.1.8. Rating items and providing feedback. A brief working definition of each ACT process was provided to the expert panel in order to facilitate their ability to rate items. The panel were also informed of the criteria by which items would be either retained or excluded from successive rounds based on their aggregated ratings. In round one, panel members were also given the option to pick an item which was: a) a poor indicator of the specific ACT process; and b) an excellent indicator of the ACT process, and explain why this was case in a 'free-text' box. This qualitative information initially helped to give us some idea about the basis on which participants were rating items. Panel members were also asked to consider whether any aspect of the particular ACT process under consideration was not being assessed by the initial item pool, and describe what was missing in a free-text box if so. Again this provided us with information about whether the panel thought that the breadth and scope of each ACT process was being adequately covered by the initial item pool. Greater scope to provide feedback about each item individually was offered in round three by providing panel members with a free-text box for comments after each item. This was not provided in the initial rounds of rating given the greater number of items in these rounds, and the possible risk of attrition if participants were required to provide more feedback. At the end of each round panel members were each asked if they could be contacted again for the next round of item rating. A tabulated summary of the items removed and panel member's feedback in relation to these and other items was provided at the beginning of the next round (see Appendix F).

2.1.9. Consensus definition and the Delphi process. Diamond et al.'s (2014) best practice criteria for reporting Delphi studies (see Table 6 below) was used as a framework during the planning of phase one of our study. Diamond et al. developed these criteria based on their systematic literature review of Delphi studies and the limitations they found across these studies, including no a priori definitions of consensus. However, where consensus was pre-specified it was operationalised in a number of ways, including: 1) a

predefined formal measure of agreement (e.g. Cronbach's alpha); 2) cut-off scores for percentage of agreement; 3) cut-off scores for percentage of disagreement; proportion of scores within a range; and 4) a decrease in variance of group responses across rounds. Based on Diamond et al.'s suggestions, I sought to use a formal metric of agreement to ascertain consensus amongst the expert panel's rating of items, and chose Cronbach's alpha (α) for this purpose. I also predefined consensus as α = ≥0.80 in our study. This threshold is also recognised as an appropriate level of reliability more generally within the psychometric literature (e.g. Cortina, 1993; Nunnally & Bernstein, 1994; Streiner, 2003).

Whilst the number of rounds of consensus building can theoretically be infinite if consensus is never reached, pragmatically three to four rounds of rating is considered sufficient to balance the need for a detailed discussion on what is being debated, whilst also considering time, cost and participant fatigue and burden. Some evidence suggests that in most Delphi studies, three rounds of rating are needed to reach a pre-determined level of consensus (Kalaian & Shah, 2006). Others recommend however that if consensus is not achieved, the Delphi process should stop when no new information is gained from further rounds of administering the items (Linstone & Turoff, 1975). Based on the possibility that consensus wouldn't be achieved, it was decided that the Delphi process should cease when there was no change or minimal change in the Cronbach's alpha between two successive rounds of rating.

Table 6

Proposed Key Methodological Criteria to Report in Publications of Delphi-Consensus Studies (adapted from Diamond et al., 2014)

	Criteria
1.Study Objective	Does the Delphi study aim to address consensus? Is the objective of the Delphi study to present results (e.g. a list or statement) reflecting the consensus of the group, or does the study aim to merely quantify the level of agreement?

2.Participants	How will participants be selected or excluded?
3.Consensus definition	How will the consensus be defined? If applicable, what threshold value will be required for the Delphi to be stopped based on the achievement of consensus? What criteria will be used to determine when to stop the Delphi in the absence of consensus?
4.Delphi process	Were items dropped? What criteria will be used to determine which items to drop? What criteria will be used to determine to stop the Delphi process or will the Delphi be run for a specific number of rounds only?

2.1.10. Results of the Delphi. In round one, a total of 46 items (43%) from the initial 106 item pool were retained for further rating based on achieving a rating of 'good' or 'excellent' by ≥70% of participants. 60 items (57%) were therefore excluded for failing to reach this threshold. Statistical Package for the Social Sciences (SPSS) software was used to perform a Cronbach's alpha analysis and identify whether there was consensus amongst panel members in terms of item ratings. Reliability analyses were performed and found the following: openness to experience and detachment from literality - acceptance (.84); openness to experience and detachment from literality - fusion (.87); self-awareness and perspective taking (.75); and motivation and activation (.80). Based on our pre-defined consensus criteria (≥.80), consensus was not achieved across all ACT processes and a second round of rating therefore took place. It is important to note that I included items in round two even from those ACT processes that achieved consensus in round one, as our intention at this stage was to further reduce the large pool of items to a select few with the best face and content validity.

11 panel members completed the survey in round two, with two panel members dropping out from the last round despite agreeing to be contacted again. 12 items (26%) from the remaining 46-item pool were excluded for failing to achieve rating thresholds. Cronbach's alpha analyses for each of the ACT processes was calculated as: openness to experience and detachment from literality - acceptance (.71); openness to experience and detachment from literality - fusion (.40); self-awareness and perspective taking (.81); and

motivation and activation (.42). This evidence suggests that the level of internal agreement amongst panel members decreased between rounds one and two across all ACT processes. However, it is well established within the psychometric literature that Cronbach's alpha values are sensitive to the number of items under analysis, and a greater number of items can artificially inflate alpha values (Cortina, 1993); the reverse is therefore also true, and thus the observed decreases in the alpha values in round two may be due to the removal of a large number of items following round one. As consensus was not achieved across all ACT processes a third round of rating took place.

In round three I introduced 11 additional items, meaning there was a total of 45 items to be rated at this stage. These items were introduced based on the expert panel's feedback about what aspects of the ACT processes they considered missing from items in the previous rounds. Most of these items were devised by the panel members, and two items were taken from Gird and Zettle's (2013) self as context scale. Additionally, changes were made to the wording of 10 existing items in order to make them more comprehensible, and/or more theoretically precise. Again, these changes were based upon the expert panel's feedback in round two. 10 panel members completed the survey in round three, with one panel member dropping out from the previous round. Seven items (15.6%) from the remaining 45-item pool were initially excluded for failing to achieve rating thresholds. Cronbach's alpha for each of the ACT processes was calculated as: openness to experience and detachment from literality - acceptance (.81); openness to experience and detachment from literality - fusion (.94); self-awareness and perspective taking (.73); motivation and activation (.87). Based on these calculations, the expert panel failed to achieve consensus only on items relating to the self-awareness and perspective taking ACT dyadic process. The analysis showed that if one self as context item was removed the alpha levels for the overall 'subscale' would increase to beyond our threshold for consensus (.84). This item was only introduced in round three following feedback from the panel that the 'self as context' aspect of this construct was missing. Moreover, the item only just reached our threshold for being retained, with 70% rating it as 'good' or 'excellent' and 30% rating the item as either 'fair' or 'neutral'. Whilst the removal of this item may have led to the

underrepresentation of this construct within our measure, it was decided that a further round of rating of these items would be problematic for the panel in terms of further devotion of time. The item was therefore removed and the Delphi process completed.

The 37 items retained from round three were all candidates for inclusion in the final measure (See Appendix G). Our initial intention was to further reduce the number of items by selecting those with the highest face and content validity rating amongst the expert panel. However, this resulted in the selection of items with high overlapping content, and thus there was a strong risk of creating an overly narrow scale of ACT processes. Moreover, the test construction literature (e.g. Clark & Watson, 1995) suggests that more items should be supplemented during the validation of a measure than one expects to include in the final measure, as this allows items which perform poorly in the validation process to be dropped and replaced with better performing items. I therefore decided to subject all 37 items retained from round three to the validation process, and then reduce these items at a later date.

2.1.11. Testing the readability of items in the measure. Whilst the ACT expert panel were asked to provide feedback on the readability of items, we sought to supplement this by testing the readability of the CompACT's initial 37-item pool using: a) the flesch reading ease score test (FRES); and b) the flesch kincaid grade Level test based on recommendations by Streiner and Norman (1995). Each of these tests was performed using free online software (available from https://readability-score.com).

The FRES rates text on a 100-point scale, with higher scores indicating that a text is easier to read. For documents intended for the general population, a FRES of between 60-80 is recommended. Our measure achieved a score of 78.4 (see Table 7 below). The Flesch-Kincaid test examines texts based on sentence length and word length, and derives a score to indicate the number of years of education necessary to comprehend the text based on an American grading system. For example, a score of '9' means a ninth grader (i.e. a student between 14-15 years old) can understand the document. Lower scores indicate the document requires fewer years in education to understand and is therefore

easier to read. The recommended score for general documents is between 7.0 - 8.0. Our measure achieved a score of 5.9 meaning that it should be understandable for someone between 10-11 years old.

In determining the format of a measure, DeVellis (1991) suggests that participants need enough response options that will permit sufficient variability in their responses, but not so many options as to increase their fatigue and boredom. Another consideration is how to facilitate participant's ability to respond discriminately between different options, which will depend in part on the wording, and on the physical depiction of these response options. Similar to a number of other ACT process measures, I chose to use a seven-point Likert scale from 0 ('strongly disagree'') to 6 ('strongly agree').

Table 7

The Reading Ease, Grade Level and Text Statistics Calculations for the CompACT

Reading Ease	
Scores range between 0 and 100. A higher score indicates easier readability. For documents intended for the general population a score between 60 and 80 is recommended.	
Readability Formula	*Score*
Flesch-Kincaid Reading Ease	78.4
Grade Levels	
A grade level (based on the USA education system) is equivalent to the number of years of education a person has had. The recommended score for general documents is between 7.0 and 8.0.	
Readability Formula	*Grade*
Flesch Kincaid Grade Level	5.9
Text Statistics	
Character Count	2163
Syllable Count	716
Word Count	532

2.2. Phase two: Testing the dimensionality, validity and reliability of the CompACT in a non-clinical sample

2.2.1. Participants. During the development of a new measure, an adequate sample size is considered necessary to avoid several issues, including: poor representation of the test sample comparative to the target population; constriction of the target attribute (i.e. the level of attribute present) in the test sample comparative to the target population; and unstable patterns of covariation amongst items, such that the reliability coefficients of each item (i.e. Cronbach's alpha levels) are artificially affected (DeVellis, 1991). Several authors suggest 300 participants is a sufficient sample for the validation of a new measure (e.g. Clark & Watson, 1995; DeVellis, 1991; Nunnally, 1978). Others suggest a sample size as low as 200 is permissible (e.g. MacCallum, Widaman, Zhang, & Hong, 1999). I erred on the side of caution and sought to recruit at least 300 participants to phase two of the study. The final number of participants totalled 377, which were recruited from a non-clinical sample.

Our choice to use a non-clinical sample was also based upon Bond et al.'s (2011) suggestion that whilst the ACT model is applicable to psychopathology, it is intended for a much broader application to human adaptability and behavioural functioning across a variety of contexts, such as job performance (e.g. Bond, Flaxman & Bunce, 2008) and sport skills (Bond, Flaxman, van Veldhoven & Biron 2010). Such contexts will also shape and determine what processes ACT interventions will target and how it will do this. For example whilst ACT interventions in psychotherapy may target a shift in the avoidance of internal experiences, in job performance settings ACT interventions may focus on identifying team values, improving problem solving, or enhancing contingency sensitivity. Therefore, psychometric measures used to determine ACT's effectiveness in both of these contexts need to be able to measure both 'pathological' and 'non pathological' processes; in ACT the processes involved in psychopathology are the same as those involved in adaptability, and thus measures assessing these processes do not need to derive nor develop from clinical samples per se.

2.2.2. Measures: Convergent and discriminant validity. Despite our earlier criticisms, the AAQ-II and the AAQ both remain the most widely used outcome measures in ACT studies (Ruiz, Langer Herrera, Luciano, Cangas & Beltran, 2013). I therefore sought to establish the CompACT's convergent validity by asking participants to complete the AAQ-II. In addition to its widespread use, the AAQ-II was also selected above other ACT process measures to ascertain the convergent validity of the CompACT, as some of the items from these other measures were incorporated into the CompACT, whilst all items from the AAQ-II failed to achieve sufficient ratings to be included within the measure. Therefore, any correlational analyses between the CompACT and these other measure would be artificially inflated by possessing the same items.

The AAQ-II is puported to measure the six ACT processes which form psychological flexibility. The measure has seven items and utilises a seven-point Likert scale (1 'never true'; 7= 'always true') requiring participants to rate how true each statement is for them (e.g. 'I'm afraid of my feelings'). Bond et al. (2011) found that the AAQ-II had good internal reliability (.78-.88) and test-retest reliability at 3 months (.81) and 12 months (.79). It was predicted that there should be a moderate to high correlation[18] between the CompACT and the AAQ-II as they both seek to measure similar theoretical constructs. However, given our suggestion that many of the items within the AAQ-II are ostensibly distress items, I did not expect nor wish for very high correlations with the CompACT, as I did not want to replicate the AAQ-II's limitations in this respect.

The CompACT's discriminant validity was assessed by asking participants to complete a measure of social desirability, as this was predicted to have a minimal relationship to the ACT processes underpinning the CompACT. The Marlowe-Crowne Social Desirability scale (MCSD; Crowne & Marlowe, 1960) is a 33-item measure which captures when a participant's responding is constrained by social desirability. Participants are asked to provide dichotomous

[18] Whilst the CompACT was intended to be a measure of psychological flexibility, the predominance of negatively valenced items within the measure means that it is actually a measure of psychological inflexibility. As we had not foreseen this prior to examining the CompACT's associations with other measures, the predicted direction of the CompACT's correlations with these measures detailed in this section of the thesis is based on the CompACT being a measure of psychological flexibility.

'yes' or 'no' responses to items, with higher scores indicating more socially desirable responses. Several authors have attempted to develop shortened versions of the MCSD. Loo and Loewen (2004) conducted confirmatory factor analyses (CFA) and item scale analyses to evaluate different versions of the short form MCSD, and to ascertain how they compare to the full MCSD. Their findings suggest that the short forms have comparable psychometric properties to the full MCSD, and Ballard's (1992) 13-item version is the best based on RMSEA and AGFI fit indexes. The internal consistency for Ballard's MCSD has been reported between .62 to .76 (Ballard, 1992; Loo & Thorpe, 2000; Zook & Sipps, 1985). Evidence also suggests that there is a strong correlation (.91-.97) between Ballard's MCSD and the full version MCSD (Fischer & Fick, 1993; Loo & Thorpe, 2000). I chose to use Ballard's short form MCSD over the full MCSD in order to prevent participant burden and fatigue associated with having to complete too many items. I predicted there should be minimal correlation between the CompACT and the short form MCSD.

2.2.3. Measures: Concurrent validity. The Depression Anxiety Stress Scale (DASS-21; Lovibond & Lovibond, 1995) is a self-report questionnaire consisting of three subscales designed to measure the severity of a range of symptoms common to depression, anxiety and stress (seven items per symptom). Participants are asked to consider their experiences over the preceding week and indicate how much each statement applies to them on a four-point Likert scale from 0 ('did not apply to me at all over the last week') to 3 ('applied to me very much or most of the time over the past week'). The measure has demonstrated good internal reliability (.82-.97) in clinical and non-clinical samples (e.g., Henry & Crawford, 2005; Lovibond & Lovibond, 1995). The DASS-21 was chosen over other more widely used distress measures (e.g. Beck Depression Inventory-II [BDI-II; Beck, Steer & Brown, 1996]; Beck Anxiety Inventory [BAI; Beck & Steer, 1990]) as it assesses multiple distress constructs within one measure (depression; anxiety; and stress); thus I highly valued its parsimony given our aim to reduce the risk of participant burden and attrition. It was predicted that higher scores on the CompACT should correlate with lower scores on the DASS-21.

The Short Form Health Survey (SF-36; Ware and Sherbourne, 1992) is a 36-item measure of health status that broadly taps into areas of physical health, sense of well-being, quality of life and ability to function. The SF-36 requires participants to indicate on a five-point Likert scale (from 'poor' to 'excellent') their view of their health across eight domains: physical functioning; role limitations due to physical problems; social functioning; bodily pain; mental health; role limitations due to emotional problems; vitality; and general health perceptions. Two summary component scores are derived from the SF-36: the Physical Component Summary Score (PCS); and the Mental Health Component Summary Score (MCS). A higher score suggests better health and quality of life. The SF-36 has been described as a 'gold-standard' for self-report assessments of general health (McCracken, Chilcot & Norton, 2015).

A shortened version of the SF-36 has been developed (SF-12v2; see Ware, Kosinski & Keller, 1996; Ware, Kosinski, Turner-Bowker & Gandek, 2002) using regression methods. Ware et al. (1996) found high correlations between the SF-12v2 and SF-36 and good test-retest reliability scores at 2-weeks for the SF-12v2 (.76-.89). The measures were also found to have good internal consistency for the physical health subscale (.92) and mental health subscale (.88). The SF-12v2 has also been found to be as responsive to longitudinal changes in health status as the SF-36 (Jenkinson et al., 1997; Maruish, 2012). We therefore used the SF-12v2 in our study and predicted that higher scores on the CompACT should correlate with higher scores on the SF-12v2 indicating a better quality of health and life functioning.

3. Extended results

3.1. Issues with multicollinearity

Extreme multicollinearity (typically observed when the correlations amongst variables are >.8) can be problematic as it implies overlap between variables and therefore redundancy in their unique contribution to measuring target constructs. However, the small magnitude of multicollinearity found amongst the CompACT's initial items is not problematic per se on several grounds: 1) empirically, none of the inter-item correlations are >.80, and only 4 of 666 pairings have a correlation >.70; 2) if multicollinearity was a major issue, the analysis wouldn't resolve, yet a stable factor structure for the CompACT was later found; 3) conceptually, a reasonably high level of multicollinearity in factor analysis is wanted, as I are assuming that the items/variables can be reduced to a smaller number of factors. Herein lies a crucial difference between multicollinearity in factor analysis and measurement development, and multicollinearity in regression. In factor analysis the overall factor scores rather than the individual variables are entered into analyses. In regression, variables are modelled as independent and are therefore entered into analyses separately. Regression analyses are thus more likely to be affected by multicollinearity between individual items compared to when this occurs in measure development; 4) the critical requirement for factor analysis is that perfect multicollinearity (i.e. items/variables that are perfectly correlated) is avoided, and this is indeed the case for the CompACT's items; and 5) whilst an

analysis of the contents of the CompACT's items reveals that there are a number of items that overlap, pragmatically this is necessary in multi-item scaling in order to: a) examine the reliability of participant's responding to items (DeVellis, 1991); and b) provide scope to develop short forms (or even parallel forms) of the CompACT containing different, but conceptually similar items.

3.2. Identification of the CompACT's factor structure

Factor analytic strategies are used to ascertain the dimensionality of a measure, when more than one latent variable is postulated to underlie a set of items (Floyd & Widaman, 1995). Ferguson and Cox (1993) suggest that factor analysis can be used to either: a) 'confirm' the pre-specified underlying factor structure of a measure by examining the 'fit' of new data to a measure (i.e. CFA); or b) it may be used in a more 'exploratory' fashion whereby the factor structure of a measure is determined by the data rather than specified a priori (i.e. EFA). Whilst I hypothesized that the CompACT's conceptual underpinnings may best be captured by three dyadic ACT processes/factors, ultimately I believed that the dimensionality of the measure should be determined by the data via an inductive versus deductive process. I therefore chose to perform an EFA. As is typical within the test development literature, future studies of the CompACT will be able to test out the number of factors underpinning the CompACT as identified by the EFA in this study, by specifying this a priori and performing a CFA with new samples.

3.2.1. Adequacy of sample size for factor analysis. There is a lack of agreement within the literature as to the number of cases needed to perform an EFA, with some suggesting that the absolute number of cases (*N*) is key (e.g. Comrey & Lee, 1992), whilst others suggest a subject-to-variable ratio (*p*) is more important (e.g. Kline, 1979). However, MacCallum et al. (1999) argue that there is little empirical support for the superiority of one set of criteria over another. The adequacy of a sample size for an EFA can however be determined statistically on a known dataset, by performing The Kaiser Meyer Olkin (KMO) Test (Kaiser, 1970) and Bartlett's Test of Sphericity. The KMO measure varies between 0 and 1, and values closer to 1 are better with a minimum value of .5 necessary for conducting a factor analysis (Kaiser, 1974).

Hutcheson and Sofroniou (1999) provide useful additional guidelines for categorising different KMO values: >.7 ('middling'); >.8 ('meritorious); and .9 ('marvellous). Bartlett's Test of Sphericity tests the null hypothesis that the correlation matrix is an identity matrix. If the test is significant this would suggest that the correlations between variables are overall significantly different from zero which is desirable, as without this there is a suggestion that every variable correlates badly with all other variables (Field, 2013).

3.2.2. Extraction method. I decided to extract the factors using Principle Axis Factoring (PAF), which is a type of EFA (Pett, Lackey & Sullivan, 2003). PAF is considered a useful extraction method for theoretically exploring the underlying factor structure of a measure (Field, 2013). Whilst PAF may produce similar results to other common extraction methods such as Principal Components Analysis (PCA), PAF also allows for error within each item of a measure unlike PCA (i.e. it accommodates for the fact that each item imperfectly captures its intended construct due to random error). PAF does this by examining the shared variance amongst variables, rather than on sources of error that are unique to individual items (Kline, 1998). Allowing for random error within each item in a measure is desirable when the goal is theory development, or when wanting to generalise the factors underlying a measure to broader populations. However, in order to be able to generalise our data, in future I must use different samples and test whether the same factor structure for the CompACT can be found (i.e. I would need to cross-validate the data).

3.2.3. Selection of rotation method. The aim of rotation is to simplify the factor structure and to create a clear distinction between the factors by maximising high loadings and minimising low loadings (Williams, Brown & Onsman, 2010). Rotation therefore allows for a clearer interpretation of how many factors may underpin a measure. Nunnally and Bernstein (1994) suggest that where factors may correlate as elements of a more superordinate factor, an oblique rotation method should be used. Conversely, where factors in an analysis are assumed not to correlate, an orthogonal rotation method is preferred. Broadly, within the social sciences factors are proposed to correlate to varying degrees, as it is difficult to partition human behaviour into unrelated

and independent units (Costello & Osborne, 2005). Whilst I predicted that a three-factor solution would be sufficient for the CompACT, ACT theory posits that these factors are interrelated, and ultimately should collapse further into the single factor of psychological flexibility. Conceptually then, an oblique rather than orthogonal rotation was indicated for use in the analysis of the CompACT, as this would allow for the hypothesized inter-relation between the measure's factors to be identified. Oblique rotation methods have been applied in the factor analysis of other ACT process measures that are conceptually underpinned by the same model of ACT processes as the CompACT (e.g., AAQ-II; BEAQ; CFQ).

It is of note that some authors suggest that the choice of rotation method has little impact on the goal of producing a simplified structure in factor analysis (Kim & Mueller, 1978), and that if factors are truly uncorrelated as assumed by orthogonal rotation methods, orthogonal and oblique rotation produce nearly identical results anyway (Costello & Osborne, 2005).

3.2.4. Criteria determining the number of factors to extract. A number of criteria have been proposed for determining how many factors to extract in a factor analysis. Recommendations suggest using multiple versus single criteria to accommodate for the weaknesses of each criterion when used alone (Williams et al., 2010). A key extraction method utilised was Parallel Analysis. This method is considered by a number of researchers to be the best and most accurate means of determining the number of factors to extract (Velicer, Eaton & Fava, 2000; Zwick & Velicer, 1986). The analysis works by generating random eigenvalues from randomly generated correlation matrices. These random eigenvalues can be then compared with the actual eigenvalues from a dataset. Where the actual eigenvalues exceed the random eigenvalues at $p < .05$, these factors are retained.

SPSS does not allow for a Parallel Analysis to be performed. I therefore used an online 'Parallel Engine Analysis' developed by Patil, Singh, Mishra and Donovan (2008; available from http://smishra.faculty.ku.edu/parallelengine.htm). The engine utilizes an SAS-based code written by O'Connor (2000). The eigenvalues generated were based on conducting a PCA rather than a PAF.

O'Connor states that whilst some researchers suggest that the eigenvalues from one extraction method (e.g. PCA) should not be used to determine the number of factors for the other extraction method (e.g. PAF), there is no conclusive evidence to support this assertion. Moreover, some evidence suggests that the use of PCA-based random eigenvalues for PAF extraction can perform as well, or even better than using PAF-based random eigenvalues (see Ruscio & Roche, 2012). PAF-based random eigenvalues are also typically small and overly cautious, arguably resulting in more factors being retained than necessary. When I examined the random eigenvalues using a PAF this was confirmed to be the case.

3.2.5. Derivation of three factor structure for the CompACT. A five-factor solution was also found to meet the mean-value criterion for a Parallel Analysis from our initial PAF. However, when a subsequent five-factor solution was run, none of the items on the fifth factor met criteria for retention (i.e. loadings of >.45 on factors; cross-loading items <.2 between the primary factor and the cross-loading factor(s); Tabachnick & Fidell, 2007). This therefore brought us back to a four-factor solution, and from there to three factors, as only a three factor structure produced a stable solution.

3.3. Choosing which items to exclude/retain

The specificity and magnitude of each item's loading onto factors they are theoretically predicted to can be used to select which items to retain during measurement development. As Kline (2014) notes, in order to obtain a simple factor structure, scale items should ideally load highly onto a single factor, and should demonstrate a low (to zero) loading onto other theoretically unrelated factors.

Numerous threshold values have been proposed to determine which items to retain. For example, Ferguson and Cox (1993) suggest that items should be dropped if they either have: 1) loadings <.4 on all factors; or 2) a loading of ≥.4 on more than one factor (i.e. cross-loading items). Tabachnick and Fidell (2007)

suggest removing items with: 1) loadings <.45 on all factors; and 2) cross-loading items with a gap of at least .2 between the primary factor and the cross-loading factor(s). However, there is little evidence to support the superiority of one set of criteria over another, and the criteria themselves are arguably subjective. Instead, the choice for selecting which criteria to employ for retaining items should be guided by what supports the clearest and simplest factor structure (Kline, 2014). I therefore successively applied all of the above criteria to see which yielded the clearest and simplest factor structure. I found Tabachnick and Fidell's criteria the most suitable for this purpose, and removed seven items from the CompACT using these criteria (see Appendix I).

4. Extended discussion

4.1. Strengths of the Delphi methodology

A potential weakness of the AAQ-II is that if authors of such measures are part of a well-established working group, there is the possibility of coalescence and a convergence of opinion (e.g. on which items to exclude and retain) due to psychological processes such as 'groupthink' (see Janis, 1972). The Delphi method I used is advantageous in helping to minimise issues such as groupthink, whilst also promoting consensus, as it is recommended that participants recruited to form the expert panel are diverse enough to represent the full breadth of opinion on an issue, but not so diverse as to risk inertia in arriving at consensus. An analysis of the professional backgrounds of the ACT experts used in our study shows that the panel was composed of researchers and clinicians employed within a variety of different services, and within different geographical locations. I therefore believe that the composition of the panel in our study was adequately diverse.

Whilst the authors of the AAQ-II also purportedly used ACT experts to screen items, there is little discussion on how this was undertaken, and little deliberation on what basis items were selected/excluded. For example, Bond et al. (2011) do not discuss the level of agreement/disagreement amongst their ACT experts regarding the face and content validity of the items they devised; nor do they provide details of the idiosyncratic opinions of their experts

regarding each item. Thus, there remain questions as to whether some experts had similar concerns about the AAQ-II's items conflating ACT processes with distress constructs as I have suggested, and to what extent this disagreement may have been. In contrast, I suggest that the use of a Delphi methodology, and the stipulation of a rating and consensus threshold criteria a priori, means our screening of items was more robust. The Delphi method's qualitative summaries of participant feedback regarding each item also means that the basis on which our ACT experts excluded/retained items is more transparent than in the AAQ-II development study.

4.2. Factors determining the composition of the CompACT's items

The initial 106-item pool in phase one contained more positively valenced than negatively valenced items relating to the ACT processes forming motivation-activation (i.e. values and committed action). Positively valenced motivation-activation items thus numerically had a greater chance of being included in the final CompACT. Interestingly, authors of the values and committed action process measures from which motivation-activation items were extracted largely sought to operationalise these processes positively. This contrasts with developers of other ACT process measures used within our study, who defined ACT processes forming openness and self-awareness dyadic processes negatively. It is possible that unlike other processes within the hexaflex, values and committed action processes are easier to define and connote positively rather than negatively.

During the Delphi phase of our study, feedback from some ACT expert panel members implied that negatively valenced items were seen as of greater relevance, and therefore easier to endorse, for a section of our target sample (i.e. clinical populations). These items were rated higher for face and content validity, and thus were retained at the expense at positively valenced items. Additionally, some ACT expert panel members also suggested that including mixed valenced items was confusing for them during rating, and thus may also

be confusing for people completing the CompACT. Other items of different valence were finally removed due to either poor loading, or excessive cross-loading onto factors they were not designed to represent.

4.3. Effects of negatively valenced items of a measure's factor structure

It is of note that positively valenced items were found to have the more deleterious effect on factor structure in both the AAQ-II and the CompACT, which resulted in the removal in all (AAQ-II) or some of these items (CompACT). This suggests that the aforementioned problems associated with mixed measures is not confined to the influence of negatively valenced items.

It may be the case that an item's valence determines its suitability for measuring ACT processes within either clinical or non-clinical samples. This was a suggestion made by some ACT experts during the Delphi phase of our study, which resulted in negatively valenced items receiving higher face and content validity ratings because of their supposed suitability for clinical samples. An interesting area for future research may therefore be to compare whether item-wording has a differential effect on participant's endorsements of items in clinical versus non-clinical samples.

Regardless of the suitability of an item's valence for particular populations, the CompACT possesses items with a mixed valence. Mixed valence items are often included within a measure to minimise extreme responding and acquiescence that can accompanying having items phrased unidirectionally (Wong, Rindfleisch, & Burroughs, 2003). However, this was not the case for the CompACT, as items were selected purely on the basis of their apparent good face and content validity as adjudged by ACT experts. Mixed measures have previously been criticised because they can engender confusion in responding, and thus for ease of completing it is arguably better to pose items in a unidirectional manner (van Sonderen et al., 2013). Again, an area for future research may be develop an alternative version of the CompACT where all items are phrased either positively or negatively, and then assess whether this affects participant's endorsements of items.

4.4. Concerns regarding the amalgamation of particular ACT processes

During the Delphi phase of our reseach, some ACT experts argued on theoretical and practical grounds against how particular ACT processes (e.g. acceptance and defusion) were combined into a single dyadic process. The arguments they proposed for a conceptual and clinical difference between these processes were understandable. Conceptually, acceptance connotes an 'attitude' of willingness to contact rather than experientially avoid or control aversive experiences, if it is in the service of one's values; whereas defusion is the process by which the context of verbal relations between the self and negative cognitions can be disrupted, thus creating some psychological distance between the self and one's ongoing negative cognitions which influence behaviour over and above direct contingencies. Moreover, clinically the ACT interventions targeting these different processes are topographically different from each other: acceptance interventions utilise both metaphor exercises to highlight the cost of unwillingness, and experiential strategies to support people to begin to actively contact aversive internal events in the service of a valued way of living; defusion exercises also use these strategies, but focus on 'deliteralising' verbal relations between the self and negative self-evaluative thoughts (Harris, 2009). Together, these points lends credence to the idea that these ACT processes may indeed be separate and distinct.

However, our findings suggest that psychometrically these processes do overlap, as items examining these different processes load sufficiently onto the same factor to indicate that they must be measuring the same theoretical construct. It is of note that the CompACT's items representing acceptance/non-acceptance and defusion/fusion were taken from ACT measures specifically designed to examine these processes in isolation from other ACT processes. Thus, the factorial overlap between these processes is unlikely to be an artefact of overlapping item content and wording. Instead, this evidence suggests that ACT processes of acceptance and defusion can be more parsimoniously explained by a single factor of 'openness', rather than as two separate factors.

Whilst psychometric evidence alone is necessary but insufficient to argue that the hexaflex has excessively overlapping and therefore redundant processes, it

has been argued conceptually that ACT processes such as acceptance and defusion are functionally related processes (Flaxman et al., 2011, p. 95). Indeed, both serve to impede non-functional rule-governed behaviour in order to enable an individual to engage withi values-based action (Snyder, Lambert & Twohig, 2011). Thus, there is both statistical and theoretical grounds for arguing the hexaflex is composed of fewer processes, and this warrants further research.

4.5. Importance of capturing values in ACT process measures

ACT suggests that experiential avoidance, fusion with negative self-evaluative thoughts, and lack of present moment awareness are not necessarily problematic per se (and in some circumstances may be beneficial in the short-term). Rather, it is when a person loses contact with those aspects of life that are most valued due to chronic experiential avoidance, fusion and lack of present moment awareness, that psychological problems develop. Thus, capturing values and committed action within items is key to ensure an ACT process measures provides the most comprehensive account of behavioural functioning according to ACT theory.

4.6. Further evidence to attest to the CompACT's discriminant validity from distress measures

Evidence of the CompACT's ability to measure ACT processes that are separate to distress constructs would be further strengthened by future research demonstrating that CompACT and the DASS-21 are better represented by two separate factors. In the development of the AAQ-II, Bond et al., 2011 argued that their measure did not overlap with the DASS (Lovibond & Lovibond, 1995) by conducting a CFA, which found that a model specifying each measure as being represented by different variables provided a better goodness of fit. However, Wolgast (2014) asserts that despite the findings of their CFA, Bond et al. do not address how the AAQ-II remains poor at distinguishing between a person's experiential avoidance as an 'approach' toward private events, versus how this approach causes them distress and reduced wellbeing. In other words, the AAQ-II still confuses process with

outcome despite the fact that ACT is only interested in targeting the former directly through its treatment programmes.

4.7. Distinguishing ACT processes from symptoms of psychological distress as operationalised in self-report measures

Bond et al. (2011) attempted to undermine the argument that the correlations between the AAQ-II and the distress measure used in their study was due to common method variance (i.e. the propensity to answer similarly to multiple scales even when there is no true correlation amongst their constructs; Fiske, 1959). They sought to do this examining whether the AAQ-II resulted in different scores between participants who were engaged in different behaviours (i.e. seeking treatment for substance abuse versus not seeking this kind of treatment). The authors suggested that the statistically significant differences found between these groups of participant's AAQ-II scores proved that the AAQ-II's predictive validity was not due to common method variance. It may be useful for future research into the CompACT to adopt a similar approach to identify whether the measures yield different scores for different behaviour sets in treatment seeking versus non-treatment seeking populations.

Hayes et al. (2012) suggest that ACT interventions aim to undermine the associations people make between ACT processes and outcome constructs such as distress. Therefore, it may also be beneficial for future research to investigate this assertion by including a sample of participants who have undergone ACT, and then examine the magnitude of any differences between their CompACT and DASS-21 scores pre-post intervention. A potential problem with this research however, is that ACT arguably involves idiosyncratic language and abstract principles, that to someone who has not been socialised to the model may find difficult to grasp. For example, the concept of self as context conveys the abstract principle that the 'self' is separate to the continuous ongoing mental phenomena, such as thoughts, that one experiences. Items relating to this concept which stand to be rated as having good face and content validity run the risk of also being criticised for possessing too much jargon to be sufficiently understandable (as was the case with such items in our study, resulting in the exclusion of these items from the CompACT).

Therefore, changes in responses to ACT measures may be an artefact of this socialisation process, rather than representing actual changes in the ACT processes of interest.

Related to this research, another research stream could involve investigating whether: a) there is an association between psychometric and behavioural measures of ACT processes following an ACT intervention; b) what magnitude, if any, is the association between such measures; c) whether the strength of associations between the different ACT measures is greater than that between the ACT measures and a distress outcome measure (which would help to establish whether psychological (in)flexibility can be distinguished from distress constructs); and d) which measure accounts for the greatest variance in measures of outcome such as distress and well being. This research could help to establish whether there is a relation between these different ACT process measures despite their different methodologies. More importantly, it could also help to ascertain if dynamic behavioural measures of ACT processes are more valid or reliable than static psychometric measures of psychological (in)flexibility as some have intimated (e.g. Hayes et al., 2012). A search of the literature suggests that some preliminary research into points a) and b) has already begun, but further investigation is needed. For example, Gratz, Rosenthal, Tull, Lejuez and Gunderson (2006) found a strong negative correlation (r= -0.76) between the AAQ and a behavioural measure of willingness to experience distress.

4.8. CompACT's associations with the SF-12v2's mental health subscale

The CompACT was also found to have a stronger association with the SF-12v2's mental health subscale, and the smaller subscales of which it is comprised of, compared to other subscales within the SF-12v2. Many of the CompACT's items have been taken from ACT process measures developed to be used with a mental health treatment population. The contents of each item are therefore more likely to reflect mental health rather than other health related constructs (e.g. physical health), and thus are more likely to overlap with items from measures of similar constructs.

4.9. Suitability of the CompACT for processes of change research

Single subject or single case experimental designs (SCEDS) have historically been the principle method to explore questions relating to changes mechanisms within psychological and psychotherapy research (for an overview see Kazdin, 2011). Unlike case studies which are largely descriptive and qualitative, SCED methodologies aim to empirically establish cause and effect relationships between proposed change processes and key outcomes at the individual level. Through replication and aggregation of these effects across several individuals, SCEDs seek to generalise these findings and make broader suppositions about human behaviour (Tate, Mcdonald, & Perdices, 2008).

Several design variations of SCEDS exist, with each serving to provide different ways of controlling for biases associated with taking repeated measures of key variables over time. However, the temporal mediation of important outcomes by change processes are typically examined using an ABAB design. In phase A, a baseline measure of the outcomes of interest are taken prior to the application of treatment. This phase therefore acts as a control condition, with the individual acting as their own control. In phase B, a 'treatment' is applied and any observed changes in outcomes between phase B and A are compared to examine the effects of treatment. The treatment is then withdrawn in the second phase A. This is an important phase, as conclusions about the causal role of interventions and change processes are strengthened if outcome measures return towards the baseline levels in the initial phase A. Following this second baseline phase, the same treatment is applied again in the final phase B. Again, where changes in the outcome measures are similar in both treatment phases, this helps to establish that any observed changes were due to treatment and change mechanisms, rather than other confounding variables.

Broadly, SCEDS and ABAB designs have been employed within behavioural research as a means to compare the effects of two different treatments within a single case (see Barlow & Hayes, 1979). Given ACT is positioned within a behavioural paradigm, SCEDS have been used in research aiming to validate ACT's processes (and the overall ACT model more generally) as important mediators of psychological distress and general health and wellbeing. As ACT

is typically delivered as an overall treatment package, SCEDS have also provided a means to establish what components of ACT treatment are most effective in mediating key outcomes (for a review of lab-based ACT component studies see Levin et al., 2012). This includes: examining the role of a cognitive defusion technique in decreasing the believability and emotional impact of distressing thoughts compared to a thought control comparison condition (Masuda et al., 2004); examining acceptance and values-based action interventions as mediators of chronic pain outcomes (Vowles & McCracken, 2008); and examining acceptance and values based action interventions as mediators of pain, depression, pain-related anxiety and disability in people with trichotillomania (Flessner, Busch, Heideman & Woods, 2008). Whilst these component studies have largely support ACT's overall model, they have either used the AAQ/AAQ-II or single ACT process measures to evidence changes in ACT processes. These findings may therefore be compromised to some extent by the aforementioned limitations of these measures, and the CompACT may therefore be a more suitable measure to use within these studies in the future.

4.10. Developing short-form measures

Short-form measures aim to retain the precision and scope of the parent measure, but with fewer items which take less time to complete. Given that there are a number of critiques within the literature around how short-forms have been developed, this research may benefit from guidance within the same literature on how to develop short-form measures (see Smith, McCarthy & Anderson, 2000).

5. Critical reflections

5.1. ACT's underpinning philosophy of science

ACT is positioned within a functional contextualist philosophy of science (Biglan & Hayes, 1997; Hayes et al., 1999). Functional contexualism is an ontological philosophy, which seeks to predict and influence psychological and behavioural events with precision, depth, and scope. It is underpinned by a 'pragmatic truth criterion' (Hayes & Brownstein, 1986), meaning that a ‘true’ analysis is simply one which proves useful for the purposes of predicting and controlling behaviour. It is not truth in an ontological sense, as from a functional contextualist perspective, objective reality is unknowable; human perceptions are limited by the bias inherent in perception itself. Additionally, in a similar vein to radical behaviourism (Skinner, 1953), functional contextualism asserts that just as the subject matter one observes is influenced by contingencies of reinforcement, so too is the observer. In this way, scientists and practitioners, cannot lay claim to an ontological truth, as what is investigated and analysed is shaped by context, which in turn influences the phenomena under investigation.

A science of psychology influenced by functional contextualism would therefore suggest that an understanding of reality and behaviour that has practical and clinical utility for the purposes in question is the best that can be hoped for. A functional contextualist would also assert that this be done with two key points in mind: 1) behaviour must always be understood in relation to the context (time, place and history) in which it occurs; and 2) the intended function of the behaviour within the specific context should established (Törneke, 2010). Clearly, a risk inherent in accepting an analysis that satisfies the functional contextualist standard of being 'pragmatically useful' is accepting any analysis, no matter how spurious. To avoid this, analyses underpinned by functional contextual theories, such as ACT, must also be buttressed by standard scientific practices of replication of effects and cross-validation of findings.

5.2. Psychometrics and classic test theory

Psychometric measures or 'tests' are largely underpinned by classic test theory (CTT; see Nunnally & Bernstein, 1994 for an overview). CTT assumes the

existence of latent traits (e.g. personality characteristics such as neuroticism), which are presupposed to guide overt behaviour. Psychometrics further assume that whilst these latent traits or constructs are not directly observable, they can be measured and quantified according to how a person responds to a set of items representing these theoretical constructs. CTT therefore implicitly adopts a correspondence criteria, and infers a causal relationship between the unobservable latent theoretical constructs and a person's overt responses to test items.

This inferred relationship influences how psychometrics numerically calculate a person's performance. How a person performs on a given day on a test ('observed score'), is weighted against both the degree of imperfection inherent within the test itself ('standard error'), and also the error associated with fluctuating human variables that can affect test performance (e.g. tiredness, hunger etc.) (Lord & Novick, 1968). The notion of inherent error within a test is important, as typically within the social sciences there is interest in using psychometrics to measure changes in latent constructs. This requires repeated administration of the same psychometric. In order to assume that any observed differences in scores occurred due to some natural (or manipulated) change in the construct of interest, the measure itself must be valid and reliable (Nunnally & Bernstein, 1994). Without this, then the observed differences may be confounded by measurement error.

CTT is underpinned by elemental realist ontological assumptions (Borsboom, Mellenbergh, & van Heerden, 2003), which suggest that the true nature of reality is knowable, and can be measured objectively by discovering the constituent parts or 'elements' of reality. This mechanistic view is considered analogous to viewing reality as a machine made up of different pieces (as per current the information processing models within cognitive psychology; Ciarrochi & Kashdan, 2013). Elemental realists therefore seek to model reality accurately in order to investigate it, for example through the use of psychometric measures and test items.

5.3. Issues associated with measuring ACT processes via classic theory methods

There are two potential issues associated with our attempts to measure ACT processes via the use of a psychometric tool informed by CTT.

Firstly, ACT's pragmatic truth criterion diverges from CTT's prioritisation of correspondence criteria. CTT implicitly assumes that a person's score on a test is identical to the theoretical construct of interest itself (Borsboom, 2006). However, a functional contextualist assumes that responses to a set of items are overt behaviours shaped by environmental contingencies, rather than indicators of unseen latent traits (Hayes, Barnes-Holmes & Wilson, 2012).

Secondly and more broadly, there is also discrepancy between ACT's underpinning philosophy of functional contextualism and CTT's philosophy of elemental realism. Whilst the former proposes that reality is unknowable, the latter argues that reality is knowable and measurable if deconstructed into its elemental parts. Although functional contextualists also seek to deconstruct reality (i.e. into antecedents, behaviours and consequences), this is done so on the pragmatic basis of being useful for predicting and controlling behaviour. No assumptions are made that this corresponds to reality by functional contextualists, whereas elemental realists do make this assumption.

For these and other reasons, some have argued that CTT approaches are poorly positioned in order to measure changes in the processes of interest to functional contextual accounts of behaviour, and that other more behavioural measures are needed (for a thorough discussion of this issue, see Hayes et al., 2012). Despite this concern, there has been a proliferation of psychometric measures of ACT processes since the inception of the therapy itself, thus implying that not all within the ACT field share reservations about this methodology. I agree with Hayes et al. (2012) that it may be useful to supplement the use of 'static' psychometrics with more 'dynamic' behavioural measures of ACT processes to provide a more complementary method of

investigation. However, I argue that this does not negate the utility of psychometric measures in examining ACT processes per se.

Wilson, Bond, Ciarrochi, Hayes and Bordieri (2010) provide some useful comments on how those wanting to develop and use psychometrics within a functional contextualist paradigm. The authors suggest that, a functional contextualist approach towards latent constructs is to assume that self-report measures capture theoretically related behaviours. These behaviours can be commonly categorised using verbal labels such as 'acceptance' or ‘defusion’, not because they correspond to a 'real' common latent construct, but rather because doing so pragmatically fulfils the purpose of predicting and controlling human behaviour. In other words, the verbal descriptions of 'acceptance' and 'defusion' are flawed proxies of constructs, which 'exist' insofar as they help explain and predict behavioural phenomena. A functional contextualist (and more specifically a RFT account of language) therefore advocates caution in treating these verbal terms as exactly corresponding the behaviours they describe. Indeed, the central thesis of ACT/RFT is that language can become problematic when the verbal labels people use adopt the properties of what is being represented (Blackledge, 2007).

I agree with the aforementioned cautionary points and have thus tried to keep them in mind during all stages of the development of the CompACT, and during the subsequent analysis and interpretations of our data.

5.4. Personal reflections on the research process

5.4.1. Choosing a research project. This research was not the first project I came to in my endeavour to find something to focus on for my thesis. Previously I had gone through several iterations of different research projects, including: looking at the efficacy of a new short term treatment for trauma; investigating the effectiveness of a well established creative writing procedure (for an overview see Pennebaker, 1997) on cancer patient's experience and processing of emotions relating to their diagnosis; and examining the development and maintenance of sexual fetishes underpinned by a behavioural and RFT conceptual framework. Whilst each of these projects ostensibly appear

unrelated, I believe they all focus to varying degrees on either advancing psychological theory, and/or providing evidence attesting to the clinical utility of an approach. Both of these facets I wanted when choosing a research project. Psychological theory fascinates and captures me. I have a real desire to understand why we behave in the ways we do, and thus any research avenue along these lines intrigues me. At the same time however, I recognise that I am training to be an applied practitioner, and thus I wanted a research project that promised to have some direct clinical utility. This desire was perhaps initially informed by my experience at undergraduate level, where one of the principal areas of frustration for me was how divorced the lab-based experimental research being undertaken by my peers and lecturers was from a real-world application.

However, whilst the aforementioned research projects satisfied my ideal balance between developing psychological theory and being clinical applicable, they all lacked a third key feature; being well formed research projects that could be undertaken within the three year time frame of the clinical psychology doctorate course. Indeed, many of these projects were perhaps more suitable for a PhD level of study, rather than an applied doctorate course that requires clinical placements to be undertaken concurrently alongside conducting research. Thus, whilst these projects appealed to me, pragmatically they were not appropriate for what I needed. As time progressed and the pressure to choose and commence with my research mounted, pragmatic considerations began to carry more importance for me when choosing a project. Dr Dawson was still maintaining his role as my principal research supervisor during this time, and after a discussion with him about my concerns, the current research was raised as a potential area for investigation. To my surprise, the more and more I learned about this research, the more I realised that what was being proposed fit all of my criteria for a desirable project; the opportunity to develop a measure with obvious clinical utility in ACT treatment; the possibility of advancing ACT theory by examining whether ACT's conceptual model is composed of fewer than six distinct processes; and crucially, a project that had clearly demarcated stages of development which could be undertaken alongside my clinical work. In short, it was a perfect match for me.

Now being towards the end of the research process, I find myself reflecting on additional reasons why I like this project. At the ACBS world conference in Berlin, 2015, my project was disseminated via a poster presentation. This garnered a significant amount of attention amongst ACT practitioners and researchers, and much more so than I could have anticipated. It seems the reason for this is that others within the ACT research and clinical fields share a similar dissatisfaction with current measures of ACT processes, and believe that these measures are hindering establishing ACT's theoretical model over and above other extant and more established therapies. Indeed, several other researchers at the conference have already begun to try and develop an alternative general measure of ACT processes to the AAQ-II, but with varying degrees of success, and none so far have published their results. There is therefore excitement and curiosity around my research project and the findings, which will hopefully lead to publication and ultimately, the CompACT being used by those within the ACT community across the world.

The longevity of this project is also another reason why I am happy that I chose it. Indeed, there are several opportunities for further research into the CompACT, including further investigation of its psychometric properties if it is to be considered as a viable alternative to the ubiquitously used AAQ-II. This is something which I would like to be involved, either directly or indirectly following the completion of my studies on the doctorate in clinical psychology course. Thus, the project promises to not be a standalone piece of research, but rather the beginnings of a whole research stream which I believe is very exciting and I feel proud of.

5.4.2. Development as a researcher. Whilst my undergraduate and masters level of studies gave me the basic grounding in conducting research, ultimately in both cases the research was heavily guided by my supervisors. However, my research experience whilst being on the doctorate has advanced my competence and confidence in being able to make decisions for myself about where to take the research, and provide a sound rationale for my decision-making. Although at times this has been an anxious process, particularly around wondering whether I have been doing the 'right' thing, ultimately all forms of separation and individuation present opportunities for growth, but also failure, and it is in this conflict where anxiety lies. Fortunately, where I have found myself particularly struggling I have always been able to fall back on the wisdom, advice and most importantly, encouragement of my supervisors which I am incredibly grateful for. During the course of this research I have also been able to get grips in a more nuanced way with both the ACT literature, and the psychometric literature. This has been an intriguing process and has highlighted the discrepancies, paradoxes and general disagreement that I suspect characterises all research if one takes the time to examine what is being said at a more finite level. In order to reflect this point I wish to give a couple examples of my observations. In the psychometric literature I was quite surprised that even within the most empirically informed and quantitatively 'pure' area of research in psychometrics, there is still a significant level of subjectivity and openness to interpretation of results. I encountered this in numerous areas, including: the different criteria for deciding how many factors to extract in a factor analysis, which was largely based on conjecture; and determining the number participants necessary to conduct a factor analysis, which again was little informed by evidence over supposition. However, my surprise about this was somewhat diminished when I reminded myself of the functional contextualist framework that underpins this research. This helped me to understand that even psychometrics, with their penchant for empiricism, are ultimately devised, delivered, and interpreted by humans who are themselves shaped by their context and history. Thus, from this perspective, nothing is outside the purview of subjectivity and influence, and even investigations like those of psychometrics cannot lay claim to an ontological truth, despite doing

so; instead they provide an account of 'reality' that is true in the sense that it helps achieve their objective of quantifying and predicting behaviour.

Within the ACT literature, I was particularly surprised by the lack of consensus on some of the fundamentals of the theory including: how many distinct processes actually underpin ACT's clinical model; what is the central mechanism of change espoused within ACT (this has shifted from 'experiential avoidance' to 'psychological flexibility'); how psychological flexibility and each ACT process it is composed of should be defined and operationalised (researchers and clinicians have defined these in various different ways); the divorce between ACT and its underpinning RFT, which appear to be developing with different trajectories (for an overview of this issue see Blackledge & Drake, 2013); and judgements on how ACT's processes of change should be measured within a functional contextual consistent manner (some advocate the use of psychometrics with caveats, others suggest these measures are poor for the purposes of ACT). There also appears a schism within the ACT world about how ACT should progress. Steven Hayes, one of the chief originators of ACT, appears to be heavily embedded within RFT research, and from my readings of his work, he tries to retain the lab-based technical precision of RFT when disseminating ACT, even within the self-help literature (see Hayes & Smith, 2005). However, others such as Russ Harris, who are first and foremost practitioners rather than researchers, are developing ACT in ways they propose as having greater clinical utility by simplifying it and stripping back as much of the jargonistic language which relates to RFT. As an example of this simplification process, Harris (2009) proposes that the ACT model can be reduced and organised into a more 'triflex' centred on three processes (which are similar to our findings): 1) being present (acceptance; defusion); 2) opening up (present moment awareness; self as context); and 3) doing what matters (values; committed action).

Both of these different avenues of progressing ACT have advantages and disadvantages. Whilst Hayes' allegiance to RFT may ensure that ACT principles are disseminated in a technically correct way, there may be a risk in employing theoretically precise but somewhat jargonistic language if this does not make

sense to people. This may be the case particularly if people have not been socialised to the ACT model sufficiently, and thus they could disengage from ACT altogether. Alternatively, Harris' attempts to simplify the ACT model may mean people are better able to grasp it and derive gains from it, but this may not accord with ACT's attempts to precisely map out the theoretical mechanisms involved in the development and maintenance of psychological distress. This may further impede attempts by researchers to support ACT's theory of psychopathology over and above other extant and more established therapies such as CBT. However, despite these differences, ultimately my research and the CompACT should have applicability to the aims of both of these camps.

Portfolio word count: 29002

Extended references

Åkerblom, S., Perrin, S., Fischer, M. R., & McCracken, L. M. (2015). The mediating role of acceptance in multidisciplinary cognitive behavioral therapy for chronic pain. *The Journal of Pain*, *16*(7), 605-615.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271-285.

Arch, J.J., Wolitzky-Taylor, K.B., Eifert, G.H., & Craske, M.G. (2012). Longitudinal treatment mediation of traditional cognitive behavioral therapy and acceptance and commitment therapy for anxiety disorders. *Behaviour Research and Therapy, 50*(7-8), 469-478.

Bach, P. A., & Moran, D. (2008). *ACT in practice: Case conceptualization in acceptance and commitment therapy*. Oakland, CA: New Harbinger Publications.

Ballard, R. (1992). Short forms of the Marlowe-Crowne social desirability scale. *Psychological Reports*, *71*(3), 1155-1160.

Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis, 12*(2)*,* 199–210.

Beck A. T., Steer R. A. (1990) *Beck Anxiety Inventory manual*. San Antonio, TX: Harcourt Brace and Company.

Biglan, A., & Hayes, S. C. (1997). Should the behavioral sciences become more pragmatic? The case for functional contextualism in research on human behavior. *Applied and Preventive Psychology*, *5*(1), 47-57.

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., ... & Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice*, *11*(3), 230-241.

Blackledge, J.T. (2007). Disrupting verbal processes: Cognitive defusion in acceptance and commitment therapy and other mindfulness-based psychotherapies. *The Psychological Record,* 57(4), 555-576.

Blackledge, J. T., & Drake, C. E. (2013). Acceptance and Commitment Therapy: Empirical and theoretical considerations. In S. Dymond, & B. Roche (Ed.), *Advances in relational frame theory: Research and application* (pp. 219-252). Oakland: Context Press.

Boeschen, L. E., Koss, M. P., Figueredo, A. J., & Coan, J. A. (2001). Experiential avoidance and posttraumatic stress disorder: A cognitive mediational model of rape recovery. *Journal of Aggression, Maltreatment & Trauma, 4*(2)*,* 211-245.

Bond, F. W., & Bunce, D. (2003). The role of acceptance and job control in mental health, job satisfaction, and work performance. *Journal of Applied Psychology*, *88*(6), 1057-1067.

Bond, F. W., Flaxman, P. E., & Bunce, D. (2008). The influence of psychological flexibility on work redesign: Mediated moderation of a work reorganization intervention. *Journal of Applied Psychology*, *93*(3), 645-654.

Bond, F. W., Flaxman, P. E., van Veldhoven, M. J. P. M., & Biron, M. (2010). The impact of psychological flexibility and acceptance and commitment therapy (ACT) on health and productivity at work. In J. Houdmont, & S. Leka (Ed.), *Contemporary Occupational Health Psychology: Global perspectives on research and practice* (pp. 296-313). New Jersey, NJ: Wiley-Blackell.

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., ... & Zettle, R. D. (2011). Preliminary psychometric properties of the acceptance and action questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, *42*(4), 676-688.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425-440.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203-219.

Bosun, P., & Modrak, V. (2014). Using the Delphi method in forecasting tourism activity. *International Letters of Social and Humanistic Sciences*, *14*, 66-72.

Boughner, E., Thornley, E., Kharlas, D., & Frewen, P. (2016). Mindfulness-Related Traits Partially Mediate the Association Between Lifetime and Childhood Trauma Exposure and PTSD and Dissociative Symptoms in a Community Sample Assessed Online. *Mindfulness*, 1-8.

Branstetter-Rost, A., Cushing, C., & Douleh, T. (2009). Personal values and pain tolerance: Does a values intervention add to acceptance?. *The Journal of Pain*, *10*(8), 887-892.

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, *84*(4), 822-848.

Brown, K. W., Ryan, R. M., & Creswell, J. D. (2007). Mindfulness: Theoretical foundations and evidence for its salutary effects. *Psychological inquiry*, *18*(4), 211-237.

Cardaciotto, L., Herbert, J. D., Forman, E. M., Moitra, E., & Farrow, V. (2008). The assessment of present-moment awareness and acceptance the Philadelphia mindfulness scale. *Assessment*, *15*(2), 204-223.

Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Multivariate Behavioral Research*, *12*(3), 289-325.

Chase, J. A., Houmanfar, R., Hayes, S. C., Ward, T. A., Vilardaga, J. P. & Follette V. (2013). Values are not just goals: Online ACT-based values training adds to goal setting in improving undergraduate college student performance. *Journal of Contextual Behavioral Science,* 3(2)*,* 79-84.

Chawla, N., & Ostafin, B. (2007). Experiential avoidance as a functional dimensional approach to psychopathology: An empirical review. *Journal of Clinical Psychology*, *63*(9), 871-890.

Ciarrochi, J., & Bilich, L. (2006). Process measures of potential relevance to ACT. *Unpublished thesis, University of Wollongong, Australia*.

Ciarrochi, J., & Kashdan, T. B. (Eds.). (2013). *Mindfulness, acceptance, and positive psychology: The seven foundations of well-being*. Oakland, CA: New Harbinger Publications.

Clark, D. A. (1995). Perceived limitations of standard cognitive therapy: A consideration of efforts to revise Beck's theory and therapy. *Journal of Cognitive Psychotherapy*, *9*(3), 153-172.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological assessment*, *7*(3), 309-319.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cortina J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104.

Costello A, B., & Osborne, J,W. (2005). Best practices in Exploratory Factor Analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.

Credé, M., Chernyshenko, O. S., Bagraim, J., & Sully, M. (2009). Contextual performance and the job satisfaction–dissatisfaction distinction: Examining artifacts and utility. *Human Performance*, *22*(3), 246-272.

Creswell, J. D. W., William, T., Taylor, S., Sherman, E., David K., Gruenewald, T.L. & Mann, T. (2005). Affirmation of personal Values buffers neuroendocrine and psychological stress responses. *Psychological Science 16*(11), 846-851.

Cribb, G., Moulds, M. L., & Carter, S. (2006). Rumination and experiential avoidance in depression. *Behaviour Change, 23*(3), 165-176.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, *52*(4), 281-302.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349.

Dahl, J., Wilson, K. G., & Nilsson, A. (2004). Acceptance and commitment therapy and the treatment of persons at risk for long-term disability resulting from stress and pain symptoms: A preliminary randomized trial. *Behavior therapy*, *35*(4), 785-801.

DeCou, C. R., Lynch, S. M., Cole, T. T., & Kaplan, S. P. (2016). Dissociation mediates the association between intimate partner violence and posttraumatic stress among treatment-seeking incarcerated women. *Journal of Trauma & Dissociation*, (just-accepted).

Deighton, J., Croudace, T., Fonagy, P., Brown, J., Patalay, P., & Wolpert, M. (2014). Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: A review of child self-report measures. *Child and Adolescent Psychiatry and Mental Health*, *8*(1), 1-14.

Dell, P. F., & O'Neil, J. A. (Eds.). (2010). *Dissociation and the dissociative disorders: DSM-V and beyond*. London: Routledge.

DeVellis, R.F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage Publications.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review. recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401-409.

Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst*, *23*(2), 239-254.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, *38*(4), 319-345.

Eifert, G. H., & Heffner, M. (2003). The effects of acceptance versus control contexts on avoidance of panic-related symptoms. *Journal of Behavior Therapy and Experimental Psychiatry*, *34*(3), 293-312.

Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A users' guide. *International Journal of Selection and Assessment*, *1*(2), 84-94.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London, UK: Sage.

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe-Crowne social desirability scale. *Educational and Psychological Measurement*, *53*(2), 417-424.

Flaxman, P. E., Blackledge, J. T., & Bond, F. W. (2010). *Acceptance and commitment therapy: Distinctive features*. East Sussex, UK: Routledge.

Flessner, C. A., Busch, A. M., Heideman, P. W., & Woods, D. W. (2008). Acceptance enhanced behavior therapy (AEBT) for trichotillomania and chronic skin picking: Exploring the effects of component sequencing. *Behavior Modification*, 32, 579–594.

Fletcher, L. & Hayes, S. C. (2005). Relational Frame Theory, Acceptance and Commitment Therapy, and a functional analytic definition of mindfulness. *Journal of Rational Emotive and Cognitive Behavioral Therapy, 23*(4), 315-336.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286-299.

Franzke, I., Wabnitz, P., & Catani, C. (2015). Dissociation as a mediator of the relationship between childhood trauma and nonsuicidal self-injury in females: A path analytic approach. *Journal of Trauma & Dissociation*, *16*(3), 286-302.

Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D., & Geller, P. A. (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modification*, *31*(6), 772-799.

Fulton, J. J., Lavender, J.M., Tull, M. T., Klein, A. S., Muehlenkamp, J. J., & Gratz, K. L. (2012). The relationship between anxiety sensitivity and disordered eating: The mediating role of experiential avoidance. *Eating Behaviors*, *13*(2), 166–169.

Gámez, W., Chmielewski, M., Kotov, R., Ruggero, C., Suzuki, N., & Watson, D. (2014). The Brief experiential avoidance questionnaire: Development and initial validation. *Psychological assessment*, *26*(1), 35-45.

Gámez, W., Chmielewski, M., Kotov, R., Ruggero, C., & Watson, D. (2011). Development of a measure of experiential avoidance: The multidimensional experiential avoidance questionnaire. *Psychological Assessment*, *23*(3), 692-713.

Gaudiano, B. A. (2009a). Öst's (2008) methodological comparison of clinical trials of acceptance and commitment therapy versus cognitive behavior therapy: Matching apples with oranges? *Behaviour Research and Therapy, 47*(12)*,* 1066-1070.

Gaudiano, B. A. (2009b) Reinventing the wheel versus avoiding past mistakes when evaluating psychotherapy outcome research: Rejoinder to Öst (2009). *Unpublished Manuscript.*

Gaudiano, B. A., Herbert, J. D. and Hayes, S. C. (2010). Is it the symptom or the relation to it? Investigating potential mediators of change in acceptance and commitment therapy for psychosis, *Behavior Therapy*, *41*(4), 543–554.

Gifford, E. V., Kohlenberg, B. S., Hayes, S. C., Antonuccio, D. O., Piasecki, M. M., Rasmussen-Hall, M. L., & Palm, K. M. (2004). Acceptance-based treatment for smoking cessation. *Behavior Therapy*, *35*(4), 689-705.

Gilbert, P. (2009). Introducing compassion-focused therapy. *Advances in Psychiatric Treatment*, *15*(3), 199-208.

Gilpin, R. (2009). The use of Theravda Buddhist practices and perspectives in mindfulness-based cognitive therapy. *Contemporary Buddhism*, 9, 227‑251.

Gird, S. R., & Zettle, R. (2013). *Developing a quantitative measure of self-as-context*. In Proceedings: 9th Annual Symposium: Graduate Research and Scholarly Projects. Wichita, KS: Wichita State University, p.47-48

Gratz, K. L., Rosenthal, M. Z., Tull, M. T., Lejuez, C. W., & Gunderson, J. G. (2006). An experimental investigation of emotion dysregulation in borderline personality disorder. *Journal of Abnormal Psychology*, *115*(4), 850-855.

Gregg, J. A., Callaghan, G. M., Hayes, S. C., & Glenn-Lawson, J. L. (2007). Improving diabetes self-management through acceptance, mindfulness, and values: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *75*(2), 336-343.

Harris, R. (2007). *The happiness trap*. Boston, MA: Trumpeter Books.

Harris, R. (2009). *ACT made simple: An easy-to-read primer on acceptance and commitment therapy*. Oakland, CA: New Harbinger Publications.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008-1015.

Hayes, S. C. (1994). Content, context, and the types of psychological acceptance. Chapter in Hayes, S. C., Jacobson, N. S., Follette, V. M. & Dougher, M. J. (Eds.), *Acceptance and change: Content and context in psychotherapy* (pp. 13-32). Reno, NV: Context Press.

Hayes, S. C. (1995). Knowing selves. *The Behavior Therapist, 18*, 94-96.

Hayes, S. C. (2004a). Acceptance and commitment therapy and the new behavior therapies: Mindfulness, acceptance, and relationship. In V. M. F. S. C. Hayes, & M. M. Linehan (Ed.), *Mindfulness and acceptance: Expanding the cognitive-behavioral tradition* (pp. 1-29). New York, NY: Guilford Press.

Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual behavioral science: Creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioral Science*, *1*(1), 1-16.

Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior relations, and a behavior analytic view of the purposes of science. *The Behavior Analyst*, *9*(2), 175-190.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, *44*(1), 1-25.

Hayes, S. C., Pistorello, J., & Levin, M. E. (2012). Acceptance and commitment therapy as a unified model of behavior change. *The Counseling Psychologist*, *40*(7), 976-1002.

Hayes, S.C., & Shenk, C. (2004) Operationalizing mindfulness without unnecessary attachments. *Clinical & Psychological Science*, 11, 249-254.

Hayes, S. C., & Smith, S. (2005). *Get out of your mind and into your life: The new acceptance and commitment therapy*. Oakland, CA: New Harbinger Publications.

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

Hayes, S. C., Strosahl, K. D., Bunting, K., Twohig, M. P., & Wilson, K. G. (2004). What is acceptance and commitment therapy? In S. C. Hayes & K. D. Strosahl (Eds.), *A practical guide to acceptance and commitment therapy.* New York, NY: Plenum Press.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. New York, NY: Guilford Press.

Hayes, S. C., Wilson, K. G., Gifford, E. V., Follette, V. M., & Strosahl, K. (1996). Experiential avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting and Clinical Psychology*, *64*(6), 1152–1168.

Hayes, S. C., Villatte, M., Levin, M., & Hildebrandt, M. (2011). Open, aware, and active: Contextual approaches as an emerging trend in the behavioral and cognitive therapies. *Annual Review of Clinical Psychology*, *7*(4), 141-168.

Healy, H., Barnes-Holmes, Y., Barnes-Holmes, D., Keogh, C., Luciano, C., & Wilson, K. (2008). An experimental test of a cognitive defusion exercise. *The Psychological Record*, *58*:623-640.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *44*(2), 227-239.

Hesser, H., Westin, V., Hayes, S.C., Andersson, G. (2009). Clients in-session acceptance and cognitive defusion behaviors in acceptance based treatment of tinnitus distress, *Behaviour Research and Therapy, 47*(9), 523-528.

Hofmann, S. G., & Asmundson, G. J. (2008). Acceptance and mindfulness-based therapy: new wave or old hat? *Clinical Psychology Review*, *28*(9), 1-16.

Hsu, C.C., & Sanford, B.A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research and Evaluation*, *10*(12), 1-8.

Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. London: Sage.

Janis, I. L. (1972). *Victims of Groupthink: A psychological study of foreign-policy decisions and fiascoes.* Boston: Houghton Mifflin.

Jenkinson, C., Layte, R., Jenkinson, D., Lawrence, K., Petersen, S., Paice, C., & Stradling, D. (1997). A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies?. *Journal of Public Health Medicine*, *19*(2), 179-186.

Kabat-Zinn, J. (1982). An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. *General Hospital Psychiatry*, *4*(1), 33-47.

Kabat-Zinn, J. (1994). *Wherever you go, there you are: mindfulness meditation in everyday life.* New York: Hyperion.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401-415.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31-36.

Kalaian, S.A., & Shah, R.M. (2006). Terminating sequential delphi survey data collection. *Practical Assessment, Research and Evaluation*, *17*(5), 1-10.

Kashdan, T. B., Goodman, F. R., Machell, K. A., Kleiman, E. M., Monfort, S. S., Ciarrochi, J., & Nezlek, J. B. (2014). A contextual approach to experiential avoidance and social anxiety: Evidence from an experimental interaction and daily interactions of people with social anxiety disorder. *Emotion*, *14*(4), 769-781.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*. Oxford University Press.

Kennedy, F., Kennerley, H., & Pearson, D. (2013). *Cognitive Behavioural Approaches to the Understanding and Treatment of Dissociation*. London: Routledge.

Kim, J. O., & Mueller, C. W. (1978). *Introduction to factor analysis: What it is and how to do it.* Beverly Hills, CA: Sage.

Kline, P. (1979). *Psychometrics and psychology*. London: Academic Press.

Kline, P. (2014). *An easy guide to factor analysis*. London: Routledge.

Lauwerier, E., Caes, L., Van Damme, S., Goubert, L., Rosseel, Y., & Crombez, G. (2014). Acceptance questionnaires in patients with chronic pain: A content analysis. *Journal of Pain*, *16*(4), 306-317.

Levin, M. E., Hildebrandt, M. J., Lillis, J., & Hayes, S. C. (2012). The impact of treatment components suggested by the psychological flexibility model: A meta-analysis of laboratory-based component studies. *Behavior Therapy*, *43*(4), 741-756.

Levitt, J. T., Brown, T. A., Orsillo, S. M., & Barlow, D. H. (2004). The effects of acceptance versus suppression of emotion on subjective and psychophysiological response to carbon dioxide challenge in patients with panic disorder. *Behavior Therapy*, *35*(4), 747-766.

Lillis, J., & Kendra, K. E. (2014). Acceptance and commitment therapy for weight control: Model, evidence, and future directions. *Journal of contextual behavioral science*, *3*(1), 1-7.

Linehan, M. M. (1993b). *Skills training manual for treating borderline personality disorder*. New York, NY: Guilford Press.

Linstone, H. A., & Turoff, M. (1975). *The Delphi method: Techniques and applications*. Boston, BO: Addison-Wesley Publishing.

Lomore, C. D., Spencer, S. J., & Holmes, J. G. (2007). The role of shared-values affirmation in enhancing the feelings of low self-esteem women about their relationships. *Self and Identity*, *6*(4), 340-360.

Longmore, R. J., & Worrell, M. (2007). Do we need to challenge thoughts in cognitive behavior therapy?. *Clinical Psychology Review*, *27*(2), 173-187.

Loo, R., & Loewen, P. (2004). Confirmatory factor analyses of scores from full and short versions of the Marlowe–Crowne Social Desirability Scale. *Journal of Applied Social Psychology*, *34*(11), 2343-2352.

Loo, R., & Thorpe, K. (2000). Confirmatory factor analyses of the full and short versions of the Marlowe-Crowne Social Desirability Scale. *The Journal of Social Psychology*, *140*(5), 628-635.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores.* Oxford, England: Addison-Wesley.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour research and therapy*, *33*(3), 335-343.

Ludwig, B. (1997). Predicting the future: Have you considered using the Delphi methodology. *Journal of Extension*, *35*(5), 1-4.

Lundgren, T., Dahl, J., & Hayes, S. C. (2008). Evaluation of mediators of change in the treatment of epilepsy with acceptance and commitment therapy. *Journal of Behavioral Medicine*, *31*(3), 225-235.

Luoma, J. B., Hayes, S. C., & Walser, R. D. (2007). *Learning ACT: An acceptance and commitment therapy skills-training manual for therapists*. Oakland, CA: New Harbinger Publications.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84-99.

Maruish, M. E. (Ed.). (2012). *User's manual for the SF-12v2 Health Survey* (3rd ed.). Lincoln, RI: QualityMetric Incorporated.

Masuda, A., Hayes, S. C., Sackett, C. F., & Twohig, M. P. (2004). Cognitive defusion and self-relevant negative thoughts: Examining the impact of a ninety year old technique. *Behaviour Research and Therapy, 42(*4), 477-485.

McCracken, L. M. (2013). Committed action: An application of the psychological flexibility model to activity patterns in chronic pain. *The Journal of Pain*, *14*(8), 828-835.

McCracken, L. M., Chilcot, J., & Norton, S. (2015). Further development in the assessment of psychological flexibility: A shortened Committed Action Questionnaire (CAQ-8). *European Journal of Pain*, *19*(5), 677-685.

McHugh, L., & Stewart, I. (2012). *The self and perspective taking: Contributions and applications from modern behavioral science*. Oakland, CA: New Harbinger Publications.

Messick, S. (1991). Psychology and methodology of response styles. In R.E. Snow, & D.E. Wiley (Ed.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach* (pp. 161-200). New Jersey, NJ: Lawrence Erlbaum Associates.

Murphy M., Black N., Lamping D., McKee C., Sanderson C., Askham J. & Marteau T. (1998). Consensus development methods and their use in clinical guideline development. *Health Technology Assessment*, *2*(3), 100-108.

Neale, J., Finch, E., Marsden, J., Mitcheson, L., Rose, D., Strang, J., ... & Wykes, T. (2014). How should we measure addiction recovery? Analysis of service provider perspectives using online Delphi groups. *Drugs: Education, Prevention and Policy*, *21*(4), 310-323.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory (3rd ed.)*. New York, NY: McGraw-Hill.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396-402.

Oh, K. H. (1974). Forecasting through hierarchical Delphi. *Unpublished Doctoral dissertation*, *The Ohio State, University, Columbus*.

Öst, L. G. (2008). Efficacy of the third wave of behavioral therapies: A systematic review and meta-analysis. *Behaviour Research and Therapy*, *46*(3), 296-321.

Öst, L. G. (2014). The efficacy of Acceptance and Commitment Therapy: An updated systematic review and meta-analysis. *Behaviour Research and Therapy*, *61*(10), 105-121.

Paez-Blarrina, M., Luciano, C., Gutiérrez-Martínez, O., Valdivia, S., Ortega, J., & Rodríguez-Valverde, M. (2008). The role of values with personal examples in altering the functions of pain: Comparison between acceptance-based and cognitive-control-based protocols. *Behaviour Research and Therapy*, *46*(1), 84-97.

Park, T., Reilly-Spong, M., & Gross, C. R. (2013). Mindfulness: A systematic review of instruments to measure an emergent patient-reported outcome (PRO). *Quality of Life Research*, *22*(10), 2639-2659.

Patil, V. H., Singh, S. N., Mishra, S., & Donovan, T. (2008). Efficient theory development and factor retention criteria: A case for abandoning the 'eigenvalue greater than one' criterion. *Journal of Business Research*, 61 (2), 162-170.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science*, *8*(3), 162-166.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage.

Pilton, M., Varese, F., Berry, K., & Bucci, S. (2015). The relationship between dissociation and voices: A systematic literature review and meta-analysis. *Clinical Psychology Review*, *40*, 138-155.

Plumb, J. C., Stewart, I., Dahl, J., & Lundgren, T. (2009). In search of meaning: Values in modern clinical behavior analysis. *The Behavior Analyst*, *32*(1), 85-103.

Powers, M. B., Zum Vorde Sive Vording, M. B., & Emmelkamp, P. G.(2009). Acceptance and commitment therapy: A metaanalytic review. *Psychotherapy and Psychosomatics*, *78*(2), 73-80.

Putnam, F. W. (1997). *Dissociation in children and adolescents: A developmental perspective*. New York, NY: Guilford Press.

Rapgay, L., & Bystrisky, A. (2009). Classical mindfulness: An introduction to its theory and practice for clinical application. *Annals of the New York Academy of Science*, 1172, 148‑162.

Raskin, R., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports, 45,* 590.

Rehfeldt, R. A., Dillen, J. E., Ziomek, M. M., & Kowalchuk, R. K. (2007). Assessing relational learning deficits in perspective-taking in children with high-functioning autism spectrum disorder. *The Psychological Record*, *57*(1), 23-47.

Ruiz, F. J. (2012). Acceptance and commitment therapy versus traditional cognitive behavioral therapy: A systematic review and meta-analysis of current empirical evidence. *International Journal of Psychology and Psychological Therapy*, *12*(3), 333-358.

Ruiz, F. J., Langer Herrera, A. I., Luciano, C., Cangas, A. . & Beltran, I. (2013). Measuring experiential avoidance and psychological inflexibility; The Spanish version of the Acceptance and Action Questionnaire - II. *Psicothema, 25*(1), 123-129.

Ruiz-Párraga, G. T., & López-Martínez, A. E. (2015). The role of experiential avoidance, resilience and pain acceptance in the adjustment of chronic back pain patients who have experienced a traumatic event: A path analysis. *Annals of Behavioural Medicine*, *49*(2), 247-257.

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*(2), 282-292.

SAMHSA's National Registry of Evidence-based Programmes and Practices (2010, July). Intervention summary: Acceptance and Commitment Therapy. Retrieved from http://nrepp.samhsa.gov/ViewIntervention.aspx?id=191

Segal, Z., Williams, M., & Teasdale, J. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. New York, NY: Guilford Press.

Shapiro, F. (1989a). Eye movement desensitization: A new treatment for post-traumatic stress disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *20*(3), 211-217.

Sidman, M. (2001) *Coercion and its fallout.* Boston, MA: Authors.

Skinner, B. F. (1953). *Science and human behavior*. New York, NY: The Free Press.

Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice Hall.

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, *17*(4), 510-550.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102-111.

Smout, M. F., Hayes, L., Atkins, P. W., Klausen, J., & Duguid, J. E. (2012). The empirically supported status of acceptance and commitment therapy: An update. *Clinical Psychologist*, *16*(3), 97-109.

Snyder, K., Lambert, J., & Twohig, M. P. (2011). Defusion: A behavior-analytic strategy for addressing private events. *Behavior analysis in practice*, *4*(2), 4-13.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99-103.

Streiner, D.L., Norman, G.R. (1995). *Health Measurement Scales: A practical guide to their development and use.* Oxford, UK: Oxford Medical Publications.

Strosahl, K. D., Hayes, S. C., Wilson, K. G., & Gifford, E. V. (2004). An ACT primer: Core therapy processes, intervention strategies, and therapist competencies. In S. C. Hayes and K. D. Strosahl (Eds.), *A practical guide to Acceptance and Commitment Therapy* (pp. 31-58). New York, NY: Springer-Verlag.

Tabachnick, B.G. and Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). New York, NY: Allyn and Bacon.

Tate, R. L., Mcdonald, S., & Perdices, M., Togher, L., Schulz, R., & Savage, S. (2008) . Rating the methodological quality of single-subject designs and n-of-1 trials: Introducing the Single-Case Experimental Design (SCED) Scale. *Neuropsychological Rehabilitation*, *18*, 385-401.

Taylor-Powell, E. (2002). *Quick tips collecting group data: Delphi technique*. University of Wisconsin, Madison, WI. Retrieved April 28, 2015 from http://www.uwex.edu/ces/pdande/resources/pdf/Tipsheet4.pdf

Tjak, J.G.L., Davis, M. L., Morina, N., Powers, M. B., Smits, J. A. J., & Emmelkamp, P. M. G. (2015). A Meta-Analysis of the efficacy of acceptance and commitment therapy for clinically relevant mental and physical health problems. *Psychotherapy and Psychosomatics*, *84*(1), 30-36.

Törneke, N. (2010). *Learning RFT: An introduction to relational frame theory and its clinical application*. Oakland, CA: New Harbinger Publications.

Törneke N., Luciano C., & Valdivia-Salas, S. (2008). Rule governed behavior and psychological problems. *International Journal of Psychology and Psychological Therapy, 8*(2), 141-156.

Twohig, M. P., Masuda, A., Varra, A. A., & Hayes, S. C. (2005). Acceptance and Commitment Therapy as a treatment for anxiety disorders. In S. M. Orsillo & L. Roemer (Eds.), *Acceptance and mindfulness-based approaches to anxiety: Conceptualization and treatment* (pp. 101-130). New York: Kluwer/Springer-Verlag.

van der Kolk B A, McFarlane A C, Weisaeth L. (1996). *Traumatic stress. The overwhelming experience on mind, body, and society*. New York, NY: Guilford Press.

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE*, *8*(7).

Varra, A.A., Hayes, S.C., Roget, N., & Fisher, G. (2008). A randomised control trial examining the effect of acceptance and commitment training on clinician willingness to use evidence-base pharmacotherapy. *Journal of Consulting and Clinical Psychology*, 76(3), 449-458.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin, E. Helmes (Ed.) *Problems and solutions in human assessment* (pp. 41-71). Massachusetts, MA: Kluwer Academic Publishers.

Vowles, K. E., & McCracken, L. M. (2008). Acceptance and values-based action in chronic pain: A study of treatment effectiveness and process. *Journal of Consulting and Clinical Psychology*, *76*(3), 397-407.

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*(3), 220-233.

Ware, J. E., Kosinski, M., Turner-Bowker, D. M., & Gandek, B. (2002). User's manual for the SF-12v2 Health Survey. *Lincoln, RI: QualityMetric Incorporated*.

Ware Jr, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, *30*(6), 473-483.

Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, *51*(1), 59-91.

White, K. G., Juhasz, J. B., & Wilson, P. J. (1973). Is man no more than this?: Evaluative bias in interspecies comparison. *Journal of the History of the Behavioral Sciences*, *9*(3), 203-212.

Wicksell, R. K., Ahlqvist, J., Bring, A., Melin, L., & Olsson, G. L. (2008). Can exposure and acceptance strategies improve functioning and life satisfaction in people with chronic pain and whiplash-associated disorders (WAD)? A randomized controlled trial. *Cognitive Behaviour Therapy*, *37*(3), 169-182.

Wicksell, R.K., Renöfält, J., Olsson, G.L., Bond, F.W., & Melin, L. (2008). Avoidance and cognitive fusion–central components in pain related disability? Development and preliminary validation of the psychological inflexibility in pain scale (PIPS). *European Journal of Pain*, *12*(4), 491–500.

Williams, L. (2007). Acceptance and commitment therapy: An example of third-wave therapy as a treatment for Australian Vietnam War veterans with posttraumatic stress disorder. *Salute*, *19*, 13-15.

Williams, B., Brown, T., & Onsman, A. (2012). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, *8*(3), 794-799.

Wilson, K. G., Bond, F., Ciarrochi, J., Hayes, S., & Bordieri, M. (2010). Reconstructing measurement theory from a contextualist perspective. Conference presentation at The Association for Contextual Behavioural Science 8th World Congress, Reno, NV, June 2010. Retrieved 10th September, 2015 from http://contextualpsychology.org/system/files/Reconstructing_measurement_theory_Assessment_presentation.pptx

Wilson, K. G., & DuFrene, T. (2009). *Mindfulness for two: An acceptance and commitment therapy approach to mindfulness in psychotherapy*. Oakland, CA: New Harbinger Publications.

Wilson, K. G., Hayes, S. C., Gregg, J., & Zettle, R. D. (2001). Psychopathology and psychotherapy. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational Frame Theory: A post-Skinnerian account of human language and cognition*. (pp. 211-237). New York: Plenum Press.

Wilson, K. G., & Murrell, A. R. (2004). Values work in Acceptance and Commitment Therapy: Setting a Course for Behavioral Treatment. In Hayes, S. C., Follette, V. M., & Linehan, M. (Eds.), *Mindfulness & Acceptance: Expanding the cognitive-behavioral tradition* (pp. 120-151). New York: Guilford Press.

Wolgast, M. (2014). What does the acceptance and action questionnaire (AAQ-II) really measure? *Behavior Therapy*, *45*(6), 831-839.

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the Material Values Scale. *Journal of Consumer Research*, *30*(1), 72-91.

Zettle, R. D. (2007). ACT for depression. *A clinician's guide to using acceptance and commitment therapy in treating depression.* Oakland, CA: New Harbinger Publications.

Zettle, R. D., Rains, J. C., & Hayes, S. C. (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.

Zima, B. T., Murphy, J. M., Scholle, S. H., Hoagwood, K. E., Sachdeva, R. C., Mangione-Smith, R., ... & Jellinek, M. (2013). National quality measures for child mental health care: Background, progress, and next steps. *Pediatrics*, *131*(1), 38-49.

Zook, I. I. A., & Sipps, GJ (1985). Cross-validation of a short form of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *41*(2), 236-238.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432-442.

APPENDICES

Appendix A

Ethical approval from University of Lincoln School of Psychology Research Ethics Committee (SOPREC)

From: Soprec [soprec@lincoln.ac.uk]

Sent: 05 November 2014 13:57

To: Francis Ashley

Dear Ashley

This is to confirm that your recent resubmission for ethical approval was approved at the panel which met on 29th October 2014.

If you would like a formal letter to confirm this, please email soprec@lincoln.ac.uk with this request.

Kind regards

SOPREC

School of Psychology Research Ethics Committee |

SOPREC

College of Social Science
University of Lincoln. Brayford Pool, Lincoln, Lincolnshire. LN6 7TS

Email – soprec@lincoln.ac.uk

Appendix B

Initial pool of 106 items extracted from existing ACT process measures with factor loadings

Draft CompACT items

All items are in order of factor loading weighting except for our own self-generated items which have no factor loadings (NM, DD and AF). All factor loadings reported are based on non-clinical samples. Six core processes have been collapsed into four processes: Acceptance; Fusion; Mindfulness; and Valued Living (terms have been changed by the authors). Hayes et al. (2011) clustered these processes into three groupings based on the overlap between pairs of processes; the name of these groupings and the processes they involve are highlighted in ***red****.*

Acceptance (Acceptance and Action Questionnaire [AAQ-II] Q1-3; Philadelphia Mindfulness Scale [PHLMS] Q4-8; Brief Experiential Avoidance Questionnaire [BEAQ] Q9-22; Self generate items Q23-31)

Openness to experience and detachment from literality (acceptance; defusion)

1. My painful memories prevent me from having a fulfilling life (.82)
2. Emotions cause problems in my life (.79)
3. I worry about not being able to control my worries and feelings (.71)
4. There are things I try not to think about (.66)
5. I tell myself that I shouldn't have certain thoughts (.65)
6. There are aspects of myself I don't want to think about (.61)
7. I try to stay busy to keep thoughts or feelings from coming (.66)
8. I try to distract myself when I feel unpleasant emotions (.54)
9. Pain always leads to suffering (.67)
10. One of my big goals is to be free from painful emotions (.67)
11. I would give up a lot not to feel bad (.67)
12. I go out of my way to avoid uncomfortable situations (.63)
13. I rarely do something if there is a chance it will upset me (.63)
14. I work hard to keep out upsetting feelings (.60)
15. If I have any doubts about doing something, I just won't do it (.57)
16. I'm quick to leave any situation that makes me feel uneasy (.55)
17. I won't do something until I absolutely have to (.51)
18. I try to put off unpleasant tasks for as long as possible (.49)
19. Fear or anxiety won't stop me from doing something important (.44)
20. It's hard for me to know what I'm feeling (.42)
21. I feel disconnected from my emotions (.37)
22. When unpleasant memories come to me, I try to put them out of my mind (.31)
23. I can experience my thoughts and feelings as they are, without trying to change them
24. I can take thoughts and feelings as they come, without attempting to control or avoid them
25. When I feel a strong emotion, I have to do something to make it go away

26. When I have reason to, I can stay in situations that bring up difficult thoughts or feelings
27. I can make room for feelings and thoughts and let them pass by
28. I try to control my thoughts to avoid particular emotions
29. I can sit with my thoughts and feelings without getting wrapped up in them
30. I am willing to fully experience whatever thoughts, feelings and sensations come up for me, without trying to change or defend against them
31. I can notice that my ability to willingly experience difficult thoughts, feelings or sensations can change at times

Fusion (Cognitive Fusion Questionnaire [CFQ] Q1-5; Believability of Anxious Feelings and Thoughts Scale [BAFT] Q6-8; Avoidance and Fusion for Youth Questionnaire [AFQ-Y] Q9-14; Drexel Defusion Scale [DDS] Q15-17; Self-generated items Q18-25)

Openness to experience and detachment from literality (acceptance; defusion)

1. I get upset with myself for having certain thoughts (.83)
2. I struggle with my thoughts (.73)
3. My thoughts cause me distress or emotional pain (.67)
4. I get very entangled in my thoughts (.65)
5. I get so caught up in my thoughts that I am unable to do the things that I most want to do (.69)
6. I could lose control when I feel anxious or afraid (.67)
7. I must do something about my anxiety or fear when it shows up (.67)
8. When I am very anxious or afraid there is a good chance that I might be dying (.63)
9. I must get rid of my worries and fears so I can have a good life (.68)
10. If my heart beats fast, there must be something wrong with me (.59)
11. I push away thoughts and feelings I don't like (.59)
12. If I breathe quickly something must be wrong (.59)
13. I am afraid of my feelings (.59)
14. The bad things I think about myself must be true (.57)
15. Thoughts of hopelessness - You are feeling sad and stuck in a difficult situation that has no obvious end in sight. You experience thoughts such as ‘‘Things will never get any better". To what extent would you normally be able to defuse from thoughts of hopelessness? (.86)
16. Thoughts of self - imagine you are having a thought such as "no one likes me". To what extent would you normally be able to *defuse* from negative thoughts about yourself? (.77)
17. Thoughts about motivation or ability - imagine you are having a thought such as "I can't do this" or " I just can't get started". To what extent would you normally be able to defuse from thoughts about motivation or ability? (.74)

18. I can choose to do something even when I have doubts
19. My thoughts can be misleading
20. I don't allow thoughts to interfere with what I want to do
21. Thoughts are just thoughts, and may not be true
22. When I have a thought I have to act on it
23. When I have a thought about myself, I don't take it literally
24. I don't have to take my thoughts seriously
25. I can observe the thoughts, feelings, sensations and memories I have come and go, as if they were clouds drifting by in the sky

Mindfulness (Mindfulness Attention Awareness Scale [MAAS] Q1-15; Self-generated item Q16-25)

Self-awareness and perspective taking (self-as-context; contact with present moment)

1. It seems I am "running on automatic" without much awareness of what I'm doing (.78)
2. I find myself doing things without paying attention (.69)
3. I rush through activities without being really attentive to them (.62)
4. I do jobs or tasks automatically, without being aware of what I'm doing (.55)
5. I find myself listening to someone with one ear, doing something else at the same time (.55)
6. I drive places on "automatic pilot" and then wonder why I went there (.51)
7. I find it difficult to stay focused on what's happening in the present (.51)
8. I snack without being aware that I'm eating (.47)
9. I could be experiencing some emotion and not be conscious of it until some time later (.46)
10. I break or spill things because of carelessness, not paying attention, or thinking of something else (.45)
11. I tend to walk quickly to get where I'm going without paying attention to what I experience along the way (.45)
12. I get so focussed on the goal I want to achieve that I lose touch with what I am doing right now to get there (.38)
13. I forget a person's name almost as soon as I've been told it for the first time (.33)
14. I find myself preoccupied with the future or the past (.28)
15. I tend not to notice feelings of physical tension or discomfort until they really grab my attention (.27)
16. I can stay focussed on things that are happening here and now
17. I can choose to pay attention to the present moment
18. I notice when my thoughts are drifting to the past or future and can refocus
19. I can deliberately focus my attention and ignore distractions

20. Although my thoughts and feelings may change, I have a sense of myself as constant
21. I am more than just the thoughts and feelings that I experience from moment to moment
22. Part of me is able to notice the things that I experience without getting caught up in them
23. My thoughts come and go independently of me
24. I can notice that the thoughts, feelings and physical sensations I have are distinct from the 'me' that is noticing these thoughts, feelings and physical sensations
25. When I have thoughts, feelings or physical sensations I can 'zoom out' from them and observe and notice them for what they are

Valued Living (Engaged Living Scale [ELS] Q1-6; The Valuing Questionnaire [VQ] Q7-12; Committed Action Questionnaire Short Form [CAQ-8] Q13-15; Self-generated items Q16-25)

Motivation and activation (values; committed action)

1. I have values that give my life more meaning (VL=.72; LF=.11)
2. I know what motivates me in life (VL=.72; LF=12)
3. I believe that I've found important values to live according to (VL=.64; LF=.04)
4. I make choices based on my own values even if it is stressful (VL=.50; LF= -.02)
5. How I behave fits in with my personal wants and desires (VL=.44; LF= -.18)
6. My values are really reflected in my behaviour (VL=.54; LF= -.07)
7. It seemed like I was just going through the motions, rather than focussing on what was important to me (.87)
8. I spent a lot of time thinking about the past or future, rather than being engaged in activities that mattered to me (.80)
9. Difficult thoughts, feelings or memories really got in the way of what I wanted to do (.79)
10. I felt like I had a purpose in life (.85)
11. I was proud about how I lived my life (.78)
12. I made progress in the areas of my life I cared most about (.75)
13. I prefer to change how I approach a goal rather than quit (.85)
14. I am able to follow my long terms plans including times when progress is slow (.80)
15. I can remain committed to my goals even when there are times that I fail to reach them (.78)
16. I can keep going with something when it’s important to me
17. I can be persistent when it matters
18. I can make changes when change is important to me

19. I can stop or let go of something when it’s not what I want
20. I won’t let doubts or worries stop me from pursuing the things that I want
21. I behave in line with my personal values
22. I undertake things that are meaningful to me, even though they unsettle me
23. I act in ways that are consistent with how I wish to live my life
24. I try to overcome the hurdles which may stop me from living my life in the way I want to
25. I recognise that no matter what has happened, every moment is an opportunity to choose whether to live my life in the way that I want and can do

Appendix C

Draft email to participants targeted for recruitment to phase one Delphi study

Dear xxxx

We would like to invite you to take part in a piece of research aiming to develop a new ACT measure that can be used to track changes in ACT processes on a regular basis.

We are specifically asking you and a select number of other experts within the ACT field to rate items representing each ACT process (on the basis of their face and content validity) until there is consensus about which items best typify each ACT process. These items will then be incorporated into a final measure which will undergo a process of validation.

It is envisaged that there will be three main rounds of item rating before consensus is achieved.

It is expected that rating of items should take no more than **20 minutes**, with each subsequent round taking less time as there will be fewer items to rate.

Should you choose to take part in this research, you will need to be available to rate items via the survey for the duration of the study which should take place across 12 weeks.

For more details about the study, please paste this link into your search engine https://eSurv.org?s=LHIOMM_71bba913 - this will take you to the survey hosting the items to be rated.

Many thanks

Ashley Francis

Trainee Clinical Psychologist
Doctorate in Clinical Psychology
Universities of Nottingham and Lincoln

Under the supervision of:

Dr David Dawson
Clinical and Research Psychologist
Academic tutor on the Doctorate in Clinical Psychology at the Universities of Nottingham and Lincoln

Dr Nima Golijani-Moghaddam
Clinical and Research Psychologist
Academic tutor on the Doctorate in Clinical Psychology at the Universities of Nottingham and Lincoln

Appendix D

A screen print showing the Likert Scale upon which Delphi participants had to rate items from different ACT process measures

'Openness to experience and detachment from literality'

40%

Please rate each item below according to how well it captures the relevant or important content of the ACT process of Acceptance, which forms part of the overall process 'Openness to experience and detachment from literality'.

	Poor	Fair	Neutral	Good	Excellent
1. My painful memories prevent me from having a fulfilling life	○	○	○	○	○
2. Emotions cause problems in my life	○	○	○	○	○
3. I worry about not being able to control my worries and feelings	○	○	○	○	○
4. There are things I try not to think about	○	○	○	○	○
5. I tell myself that I shouldn't have certain thoughts	○	○	○	○	○

Appendix E

Example summary of both the items removed, and participant's feedback in relation to items in Round one of the Delphi study

Delphi round one ratings: A summary of items excluded

Following the first round of rating by the Delphi expert panel, a total of 46 items (43%) from the initial 106 item pool were retained for further rating based on these items achieving a pre-defined threshold rating of 'good' or 'excellent' by ≥70% participants. 59 items (57%) were therefore excluded as they failed to reach this threshold.

Please see Table E1 below for a list of the items excluded from the next round of Delphi panel rating.

Table 1

A list of items excluded from round 2 of rating by the Delphi expert panel for failing to reach a predefined threshold criteria in Round

ACT Process	**Item removed**	**Percentage rating as 'good' or 'excellent' (to nearest %) (items were excluded if either rated as 'poor', 'fair' or 'neutral', or rated 'good' or 'excellent' by ≤70% of participants)**	**ACT process measure item belongs to**
Openness to experience and detachment from literality -Acceptance 16/31 items excluded from Delphi round 2 rating	1.My painful memories prevent me from having a fulfilling life	61%	AAQ-II
	2.Emotions cause problems in my life	36%	AAQ-II
	3.I worry about not being able to control	61%	AAQ-II
	4.There are things I try not to think about	61%	PMS
	6.There are aspects of myself I don't want to think about	61%	PMS
	9.Pain always leads to suffering	28%	BEAQ
	11.I would give up a lot not to feel bad	55%	BEAQ
	15.If I have any doubts about doing something, I just won't do it	28%	BEAQ
	16.I'm quick to leave any situation that makes me feel uneasy	56%	BEAQ
	17.I won't do something until I absolutely have to	11%	BEAQ
	18.I try to put off unpleasant tasks for as long as possible	38%	BEAQ
	19.Fear or anxiety won't stop me from doing something important	69%	BEAQ
	20.It's hard for me to know what I'm feeling	15%	BEAQ
	21.I feel disconnected from my emotions	31%	BEAQ
	29.I can sit with my thoughts and feelings without getting wrapped up in them	61%	AF item
	31.I can notice that my ability to willingly experience difficult thoughts, feelings or sensations can change at times	38%	AF item

Delphi round 1 ratings: A summary of participant's feedback about items

The following comments were provided by the expert panel in relation to items reflecting **'Openness to experience and detachment from literality' - Acceptance**:

4) Please pick out a 'bad' item (e.g. an item with a 'poor' rating), and explain what makes it a poor indicator of the process of Acceptance.

ID	Email	First Name	Last Name	Responses (11)	View
11115304				Pain always leads to suffering, is I think, to tied up in ACT parlance and specific meanings, that will only become apparent to the client once they are in therapy. The concept is very relevant, but I would revisit the wording. Conceptually item 17 is confused I think. Many of us have to do things we would rather not and we sometimes avoid them or approach them. The appetitive/avoidance dichotomy is not necessarly an ACT one. For example, I really dislike attneding religious services of any kind and my 'have to' criteria would be if it was a wedding/funderal for a loved one; my values take over at this point. While technically I don't 'have to' and I am choosing to go inlione with my values, the distinction is too fuzzy and I think that you may get a variety of answers for different reasons	View
11115040				31 - The subject of the item is change rather than acceptance 15 - the main focus is on cautious behaviour 20 - emotion recognition is tangential to acceptance 26 - this is mainly willingness/commitment	View
11110328				#19. This item description falls moreso approaches committed action not acceptance.	View
11109861				There are things I try not to think about. This item is too general. For instance, I try not to think about if I will be lucky in some project or not, just work as best as I can in it. That would't be an example of experiential avoidance	View
11103601				17. Even thou procrastination is a common example of avoidance the item doesnt specify that the reasson of procrastination.	View
11102383				I categorised 'it's hard to know what I'm feeling' poorly because it is often hard to know what one feels and that may or may not pertain to acceptance, but it does not clearly specify lack of acceptance.	View
11101547				Item 13. This item doesn't have the important component of "something that is important to me". We wouldn't expect people to do things that were liekly to make them upset, IF those things aren't important to them in some way.	View
11100166				9 Pain always leads to suffering - its not self identified and may pull for a more generic or intellectualised response rather than measuring my own acceptance	View
11099788				"I can notice that my ability to willingly experience difficult thoughts, feelings or sensations can change at times" Wordy and awkwardly phrased.	View
11082591				9. Pain always leads to suffering Tricky on short notice to agree on what is meant by "pain" and "suffering". And questionable what the answer predicts.	View
11079950				I would strongly suggest shying away from items that have built in dysfunction (e.g., #2 "emotions cause problems in my life" and #1,3) because they blur the distinction with items that tap neuroticism and/or quality of life.	View

5) Please tell us if there are any aspects of the process of Acceptance that are not covered by the items above and if so, what aspects of Acceptance are missing.

ID	Email	First Name	Last Name	Responses (7)	View
11115304				no	View
11110328				I'd recommend you expand on the concept of willingness and what it might look like in the midst of an unpleasant internal/external experience.	View
11109861				In general, I think most items should specify that the person is feeling or thinking something and he/she is able to get entagled and avoid that experience, and perform effectively according to his/her values in their presence	View
11103601				If difficult experiences can be acknowledge as valid.	View
11100166				It should be clearer in your instruction set that these items are worded as acceptance and non acceptance You may wish to broaden out the range of internal experiences that people can be wiling to have or struggle against - memories, sensations, urges, as well as thoughts and feelings	View
11099788				Can't think of anything in particular that is missing.	View
11082591				There were some questions I took for granted are reversed.	View

Appendix F

Participant information sheet during the validation of the CompACT in phase two of our study

Trent Doctorate
in
Clinical Psychology

Participant Information Sheet

Title of Study: Investigating the validity and reliability of the Complete Assessment of Acceptance and Commitment Therapy processes (CompACT)

Name of Researcher: Ashley William Francis

Dear participant,

This research is being organised by the Universities of Nottingham and Lincoln's Doctorate course in Clinical Psychology.

Before you decide to take part in our research I would like you to understand why the research is being done and what it would involve for you. Please see information below.

What is the purpose of the study?

Acceptance has been found to be a key factor in determining the impact of life events on our psychological well-being. Specific therapies have developed to increase acceptance (e.g. Acceptance and Commitment Therapy [ACT]). A precise measure of acceptance is needed to see if ACT therapy has been effective. Currently, there are several scales which can measure a person's acceptance, but they all have limitations. We have developed the CompACT measure to address these limitations.

In order for the CompACT to be used, evidence is needed to show that the measure measures what it is intended to (i.e. acceptance) and not something

else (e.g. depression), and that is can measure this consistently and accurately across different people at different times.

What will happen if I decide to take part?

Follow the web-link address provided. This will direct you to a survey which hosts the CompACT measure and other measures which you will be asked to complete. These other measures assess depression and quality of life. This should take 20 minutes.

Do I have to take part?

Your involvement in the study is voluntary and your data will be anonymised so that we will not know who you are. You can withdraw at any time.

Will my taking part in the study be kept confidential?

The privacy of each participant will be maintained in accordance with the 'data protection principles' as per the Data Protection Act (1998). Each participant will be assigned a randomly generated number using software packages and this will serve as a reference from which participants will be referred to in any dissemination of the findings, in order to maintain anonymity.

All data will be analysed on a password protected laptop. All written documents pertaining to the analysis and write-up of the study will also be electronically password-protected and only accessible by me and my supervisors.

Expenses and payments

In order to incentivise participants to take part in the study cash prizes to the value of £100 (1st prize) and £50 (2nd prize; two available) will be offered. A winner will be selected at random and contacted once the recruitment phase is completed and participant data has been obtained.

What are the possible disadvantages and risks of taking part?

It is not foreseen that any risks or disadvantages should come about as a result of taking part in the study. Please contact your General Practitioner (GP) if after completing the measures you feel low in mood, anxious or any other symptoms which are inconsistent with how you have been feeling recently. Contact details for support organisations will also be provided at the end of the survey.

What are the possible benefits of taking part?

I cannot promise the study will be of help to you, but your participation will help us to develop a more accurate measure of therapy. This will be of benefit to people receiving psychological treatment.

What happens when the research study stops?

After completing the questionnaire, I will examine the results of all completed questionnaires and will write up the findings as a partial fulfilment of the requirements of the Doctorate in Clinical Psychology (DClinPsy) at the Universities of Nottingham and Lincoln. It is also intended that the results will be written up for publication in a scientific journal.

What if there is a problem?

If you have a concern about any aspect of this study, you should contact me via the details given at the end of this information sheet. If you remain unhappy you may wish to contact my supervisors whose details are also given at the end of this information sheet.

Who has reviewed the study?

The research is looked at by an independent group of people, called a Research Ethics Committee who belong to the University of Lincoln. The Ethics Committee are involved to ensure that your interests are protected. This study has been reviewed and given favourable opinion by the University of Lincoln Research Ethics Committee.

Further information and contact details

Principle Investigator

Ashley Francis,
Division of Psychiatry and Applied Psychology,
School of Medicine,
YANG Fujia Building,
The University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB
Email:lwxaf7@nottingham.ac.uk

Supervisors

Dr David Dawson
Dr Nima Golijani-Moghaddam
University of Lincoln,
Health,
Life and Social Sciences,
Brayford Pool,
Lincoln,LN6 7TS
Email: ddawson@lincoln.ac.uk, nmoghaddam@lincoln.ac.uk
Tel: 01522 886029

Appendix G

The CompACT's 37 items meeting threshold criteria for being retained (i.e. a rating of 'good' or 'excellent' by ≥70% of participants) following consensus being reached by participants in Round three of the Delphi study

Final 37-item pool for the CompACT

1. I tell myself that I shouldn't have certain thoughts
2. I try to stay busy to keep thoughts or feelings from coming
3. I try to distract myself when I feel unpleasant emotions
4. One of my big goals is to be free from painful emotions
5. I go out of my way to avoid situations that might bring difficult thoughts, feelings, or sensations
6. Even when something is important to me, I'll rarely do it if there is a chance it will upset me
7. I work hard to keep out upsetting feelings
8. When unpleasant memories come to me, I try to put them out of my mind
9. I can experience my thoughts and feelings as they are, without trying to change them
10. I can take thoughts and feelings as they come, without attempting to control or avoid them
11. When I have reason to, I can stay in situations that bring up difficult thoughts or feelings
12. I am willing to fully experience whatever thoughts, feelings and sensations come up for me, without trying to change or defend against them
13. I sometimes avoid things that I care about so that I can stay in control of my emotions, urges, and impulses
14. I get so caught up in my thoughts that I am unable to do the things that I most want to do
15. I don't allow thoughts to interfere with what I want to do
16. Thoughts are just thoughts – they don't control what I do
17. When I have a thought about myself, I don't have to take it literally
18. It seems I am "running on automatic" without much awareness of what I'm doing
19. Even when doing the things that matter to me, I find myself doing them without paying attention
20. I rush through meaningful/important activities without being really attentive to them
21. I do jobs or tasks automatically, without being aware of what I'm doing
22. I find it difficult to stay focused on what's happening in the present
23. I find myself so preoccupied with the future or the past that I miss out on things happening now
24. I notice when my thoughts are drifting to the past or future and can refocus
25. Part of me is able to notice the things that I experience without getting caught up in them
26. I have values that give my life more meaning

27. I have important values I choose to live by
28. I make choices based on what is important to me, even if it is stressful
29. My values are really reflected in my behaviour
30. I am able to follow my long terms plans including times when progress is slow
31. I can keep going with something when it’s important to me
32. I behave in line with my personal values
33. I undertake things that are meaningful to me, even when I find it hard to do so
34. I act in ways that are consistent with how I wish to live my life
35. I can identify the things that really matter to me in life and pursue them
36. I know what matters to me and where I want to end up in life
37. I find it hard to know what direction to take in my life

Appendix H

A list of all measures used to assess the different aspects of the CompACT's validity*

* The Short Form Healthy Survey (SF-12v2; Ware et al., 1996; 2002) cannot be included within the appendix due to copyright licencing.

Acceptance and Action Questionnaire (AAQ-II; Bond et al., 2011)

AAQ-II

Below you will find a list of statements. Please rate how true each statement is for you by circling a number next to it. Use the scale below to make your choice.

1	2	3	4	5	6	7
never true	very seldom true	seldom true	sometimes true	frequently true	almost always true	always true

1. My painful experiences and memories make it difficult for me to live a life that I would value.	1	2	3	4	5	6	7
2. I'm afraid of my feelings.	1	2	3	4	5	6	7
3. I worry about not being able to control my worries and feelings.	1	2	3	4	5	6	7
4. My painful memories prevent me from having a fulfilling life.	1	2	3	4	5	6	7
5. Emotions cause problems in my life.	1	2	3	4	5	6	7
6. It seems like most people are handling their lives better than I am.	1	2	3	4	5	6	7
7. Worries get in the way of my success.	1	2	3	4	5	6	7

Marlowe-Crowne Social Desirability Scale Short Form (Ballard, 1992)

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide if the statement is True or False for you.

	True	False
1. It is sometimes hard for me to go on with my work if I am not encouraged.	○	○
2. I sometimes feel resentful when I don't get my way.	○	○
3. On a few occasions, I have given up doing something because I thought too little of my ability.	○	○
4. There have been times when I felt like rebelling against people in authority even though I knew they were right.	○	○
5. No matter who I'm talking to, I'm always a good listener.	○	○
	True	False
6. There have been occasions when I took advantage of someone.	○	○
7. I'm always willing to admit it when I make a mistake.	○	○
8. I sometimes try to get even rather than forgive and forget.	○	○
9. I am always courteous, even to people who are disagreeable.	○	○
10. I have never been irked when people expressed ideas very different from my own.	○	○
	True	False
11. There have been times when I was quite jealous of the good fortune of others.	○	○
12. I am sometimes irritated by people who ask favors of me.	○	○
13. I have never deliberately said something that hurt someone's feelings.	○	○

Depression Anxiety Stress Scale (DASS-21; Henry & Crawford, 2005)

Please read each statement and circle a number 0, 1, 2 or 3 which indicates how much the statement applied to you over the past week. There are no right or wrong answers. Do not spend too much time on any statement. *The rating scale is as follows:*

Did not apply to me at all - 0
Applied to me to some degree, or some of the time - 1
Applied to me to a considerable degree, or a good part of time - 2
Applied to me very much, or most of the time - 3

1. I found it hard to wind down 0 1 2 3
2. I was aware of dryness of my mouth 0 1 2 3
3. I couldn't seem to experience any positive feeling at all 0 1 2 3
4. I experienced breathing difficulty (eg, excessively rapid breathing, breathlessness in the absence of physical exertion) 0 1 2 3
5. I found it difficult to work up the initiative to do things 0 1 2 3
6. I tended to over-react to situations 0 1 2 3
7. I experienced trembling (eg, in the hands) 0 1 2 3
8. I felt that I was using a lot of nervous energy 0 1 2 3
9. I was worried about situations in which I might panic and make a fool of myself 0 1 2 3
10. I felt that I had nothing to look forward to 0 1 2 3
11. I found myself getting agitated 0 1 2 3
12. I found it difficult to relax 0 1 2 3
13. I felt down-hearted and blue 0 1 2 3
14. I was intolerant of anything that kept me from getting on with what I was doing 0 1 2 3
15. I felt I was close to panic 0 1 2 3
16. I was unable to become enthusiastic about anything 0 1 2 3
17. I felt I wasn't worth much as a person 0 1 2 3
18. I felt that I was rather touchy 0 1 2 3
19. I was aware of the action of my heart in the absence of physical exertion (e.g,sense of heart rate increase, heart missing a beat) 0 1 2 3
20. I felt scared without any good reason 0 1 2 3
21. I felt that life was meaningless 0 1 2 3

Appendix I

An initial Principle Axis Factoring analysis on the CompACT's initial 37 items (N= 352) with an initial four-factor solution

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	11.331	30.625	30.625	10.840	29.296	29.296	7.857
2	4.787	12.939	43.565	4.320	11.677	40.973	7.458
3	2.444	6.604	50.169	2.027	5.479	46.452	5.951
4	1.659	4.482	54.651	1.164	3.147	49.599	5.084
5	1.463	3.953	58.604				
6	1.376	3.718	62.322				
7	1.035	2.797	65.119				

Extraction method: Principle Axis Factoring.

a. When factors are correlated, sum of squared loadings cannot be added to obtain a total variance.

Appendix J

The seven items removed from the CompACT due to insufficient factor loadings based on Tabachnick and Fidell's (2007) (loadings <.45 on all factors; cross-loading >.2 on more than one factor)

Pattern Matrix

Note. Red font highlights items/variables to be removed due to low primary loadings [<.45] or high cross-loadings [secondary loading within .20 of primary loading]).

	Factor		
	1	2	3
CompQ1Acc	.657		
CompQ2Acc	.661		
CompQ3Acc	.599	-.252	
CompQ4Acc	.737		
CompQ5Acc	.684		
CompQ6Acc	.585		
CompQ7Acc	.788		
CompQ8Acc	.555		
CompQ9Acc	.621		
CompQ10rAcc	.767		
CompQ11rAcc	.426	.294	
CompQ12rAcc	.741		
CompQ13Acc	.391		
CompQ14Def	.455		
CompQ15rDef	.503		
CompQ16rDef	.583		
CompQ17rDef	.412		
CompQ18Mind			.641
CompQ19Mind			.892
CompQ20MInd			.780
CompQ21Mind			.734
CompQ22Mind			.583
CompQ23Mind	.365		.473
CompQ24rMind		.317	
CompQ25rMind	.308	.361	
CompQ26rMot		.573	
CompQ27rMot		.688	
CompQ28rMot		.707	
CompQ29rMot		.755	
CompQ30rMot		.703	
CompQ31rMot		.701	
CompQ32rMot		.740	
CompQ33rMot		.739	
CompQ34rMot		.690	
CompQ35rMot		.734	
CompQ36rMot		.601	
CompQ37Mot		.339	

Appendix K

Stable factor loadings of CompACT's 30 items following a Principle Axis Factoring analysis, specifying a three factor solution and using a direct Oblimin rotation method

Pattern Matrix

	Factor		
	1	2	3
CompQ1Acc	**.678**		
CompQ2Acc	**.676**		
CompQ3Acc	**.620**		
CompQ4Acc	**.743**		
CompQ5Acc	**.673**		
CompQ6Acc	**.564**		
CompQ7Acc	**.783**		
CompQ8Acc	**.560**		
CompQ9Acc	**.619**		
CompQ10rAcc	**.764**		
CompQ12rAcc	**.726**		
CompQ14Def	**.461**		
CompQ15rDef	**.501**		
CompQ16rDef	**.569**		
CompQ18Mind			**.640**
CompQ19Mind			**.889**
CompQ20MInd			**.771**
CompQ21Mind			**.742**
CompQ22Mind			**.527**
CompQ26rMot		**.559**	
CompQ27rMot		**.693**	
CompQ28rMot		**.698**	
CompQ29rMot		**.763**	
CompQ30rMot		**.698**	
CompQ31rMot		**.705**	
CompQ32rMot		**.750**	

CompQ33rMot		**.733**	
CompQ34rMot		**.701**	
CompQ35rMot		**.737**	
CompQ36rMot		**.592**	

Note. All items loading >.45, and no cross-loadings within .20 of a primary loading, thus satisfying Tabachnick and Fidell's (2001) criteria for retaining items

Extraction Method: Principal Axis Factoring.

Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 5 iterations.

POSTER

Developing a new Comprehensive measure of ACT processes (CompACT)

Ashley Francis[1], David Dawson, Nima Moghaddam
Trent Doctorate in Clinical Psychology
[1]Email: lwxaf7@nottingham.ac.uk

Introduction

Valid and reliable measures of Acceptance and Commitment Therapy (ACT) processes are needed in order to establish ACT's theorised mechanisms of change (i.e. 'psychological flexibility' and the six processes of which this is composed of: acceptance; defusion; present moment awareness; self as context; committed action; and values) over and above other extant therapies.

However, the most widely used general ACT processes measure, the Acceptance and Action Questionnaire; AAQ-II[1]), has been criticised by several authors[2,3] principally in terms of its overlap and conflation with distress outcome measures that are theoretically distinct from ACT processes. At the same time, there has been a number of measures of single ACT processes developed, which arguably do not help substantiate ACT's overall model despite their utility in examining changes in these single processes. Conceptually, there are also questions as to whether there are fewer than six distinct processes underpinning ACT and extant ACT process measures.

Research Aims

Our aims were to: 1) develop a new Comprehensive measure of ACT processes (CompACT) whilst taking into consideration the limitations of current ACT process measures; 2) test out our measure's validity and reliability by examining its performance in relation to measures of related/distinct theoretical constructs; and 3) establish whether the measure's factor structure accords with Hayes et al.'s[5] clustering of ACT's six processes into three dyadic processes (see Figure 1).

Method and Results

Phase 1. Measurement development

Using a Delphi-consensus methodology, ACT experts (n=10) rated the face and content validity of 106 items. Items were taken from existing ACT process measures and two mindfulness measures, as well as generated by ourselves and the ACT experts within our study. Items which failed to reach pre-defined ratings thresholds were removed and remaining items were re-rated until a pre-defined consensus threshold was reached. After three rounds, an initial 37-item version of the CompACT was produced. None of the items extracted from the AAQ-II were adjudged by ACT experts to have sufficient face and content validity to warrant inclusion within the CompACT.

Phase 2. Investigating factor structure, validity and reliability of CompACT

The CompACT was disseminated to a non-clinical sample of participants (N= 377) mainly via social media. Participants also completed other measures to assess whether the CompACT: 1) converged with measures of theoretically related constructs (AAQ-II); 2) could be discriminated from measures of theoretically divergent constructs (Short form Marlowe-Crowne Social Desirability scale[6]; and 3) correlated with outcome variables including distress[7] and health and wellbeing[8] in expected ways. Based on ACT theory, we predicted that higher levels of psychological inflexibility should be associated with higher levels of psychological distress and lower levels of health and wellbeing.

An exploratory factor analysis found a three-factor structure for a 30-itemed version of the CompACT. An examination of the distribution and contents of item loadings also confirmed that the CompACT's three factors could be defined in accordance with Hayes et al's. three dyadic ACT processes. The CompACT performed in theoretically predicted ways against other measures, and demonstrated associations with these other measures with a similar order of magnitude to the AAQ-II, with the exception of a distress measure, where the AAQ-II exhibited a stronger correlation as expected given its purported conflation with such measures. The CompACT's associations with these measures supported ACT's central hypothesis: higher levels of psychological inflexibility were associated with higher levels of distress and lower levels of health and wellbeing.

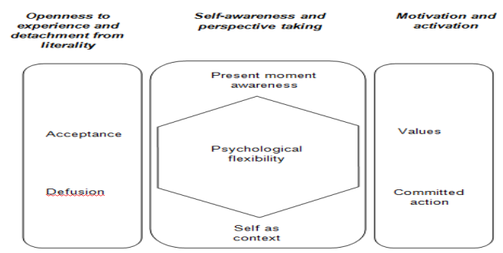


Figure 1. Hayes et al.'s (2011) three dyadic processes which underpin psychological flexibility.

Discussion

Overall, the CompACT shows initial promise as a comprehensive measure of ACT processes, and more generally as a measure of psychological flexibility. The psychometrically viability of Hayes et al.'s (2011) three dyadic ACT processes as a conceptual model for the CompACT further strengthens previous suggestions that ACT's clinical model may be composed of fewer than six distinct processes. Given that the CompACT was adjudged by ACT experts to possess items with better face and content validity than the AAQ-II, the CompACT may be considered a more precise measure of ACT processes in this regard. However, further testing of the CompACT's psychometric properties with additional samples is needed, if the measure is to be considered a viable alternative measure to the AAQ-II.

References

1. Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., ... & Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, *42*(4), 676-688.

2. Wolgast, M. (2014). What Does the Acceptance and Action Questionnaire (AAQ-II) Really Measure?. *Behavior therapy*, *45*(6), 831-839.

3. Gámez, W., Chmielewski, M., Kotov, R., Ruggero, C., & Watson, D. (2011). Development of a measure of experiential avoidance: The Multidimensional Experiential Avoidance Questionnaire. *Psychological Assessment*, *23*(3), 692-713.

4. Blackledge, J. T., & Drake, C. E. (2013). Acceptance and Commitment Therapy: Empirical and theoretical considerations. In S. Dymond, & B. Roche (Ed.), *Advances in relational frame theory: Research and application* (pp. 219-252). Oakland: Context Press.

5. Hayes, S. C., Villatte, M., Levin, M., & Hildebrandt, M. (2011). Open, aware, and active: contextual approaches as an emerging trend in the behavioral and cognitive therapies. *Annual Review of Clinical Psychology*, *7*,1), 141-168.

6. Ballard, R. (1992). Short forms of the Marlowe-Crowne social desirability scale. *Psychological Reports*, *71*(3), 1155-1160.

7. Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *44*(2), 227-239.

8. Ware, J. E., Kosinski, M., Turner-Bowker, D. M., & Gandek, B. (2002). User's manual for the SF-12v2 Health Survey. *Lincoln, RI: QualityMetric Incorporated.*