

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

School of Psychology

Why we need a better measure of acceptance:

Development and initial validation of the Southampton Acceptance Scale

by

Zoe M^cAndrews

BSc, PGCert, MSc

Thesis for the degree of Doctor of Clinical Psychology

May 2018

Word count: 20,492

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

School of Psychology

Thesis for the degree of Doctor of Clinical Psychology

Why we need a better measure of acceptance:

Development and initial validation of the Southampton Acceptance Scale

Zoe M^cAndrews

Acceptance is an important construct across models for understanding psychological distress. Several measures have been designed to capture acceptance, however, there is a lack of evidence regarding the most suitable tool. A systematic review evaluated 20 articles, reporting 32 studies, examining acceptance questionnaires. The methodological quality of included studies were evaluated using the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist. The quality of measurement properties were evaluated using criteria suggested by Terwee et al. (2007). All studies were independently reviewed by two raters, and inter-rater reliability was high. Nine instruments were identified: two unidimensional scales of acceptance, four mindfulness tools with an acceptance subscale, and three emotion regulation scales with an acceptance-based subscale. None of the measures evaluated can be recommended as having superior psychometric properties. Further research is required to demonstrate the psychometric properties of existing measures, given their significant role in evaluating acceptance-based interventions across clinical and research settings.

The lack of a valid and reliable measure of acceptance prevents researchers from drawing conclusions about the efficacy of acceptance-based interventions and identifying the role of acceptance processes in clinical change. Given that there is no current gold standard assessment tool for measuring acceptance, the present study sought to develop and evaluate a new instrument to meet this need. Across three separate studies an initial item pool was evaluated and refined. The resultant measure, named the Southampton Acceptance Scale (SAS), was then evaluated with regards to the factor structure, reliability, and validity. An initial exploratory approach was employed as a unique pool of items with many potential relationships was under investigation. These analyses were subsequently followed up with confirmatory approaches. The SAS, comprising 18-items, was shown to have a theoretically coherent two-factor structure which was validated in an independent sample. The scale has excellent internal consistency, and demonstrated convergent, concurrent and discriminant validity. The SAS has strong psychometric properties and is a promising new measure of acceptance.

Table of Contents

Table of Contents	i
Table of Tables	vii
Table of Figures	ix
Academic Thesis: Declaration Of Authorship	xi
Acknowledgements	xiii
Chapter 1 Psychometric Properties of Acceptance Measures: A Systematic Review .	1
1.1 Introduction.....	1
1.1.1 Emotion Regulation and Acceptance	1
1.1.2 Acceptance across therapies	4
1.1.3 Measurement of Acceptance.....	5
1.1.4 Objectives.....	6
1.2 Method	7
1.2.1 Search Strategy	7
1.2.2 Eligibility Criteria	8
1.2.3 Data extraction.....	9
1.2.4 Methodological Quality.....	9
1.2.5 Measurement Property Quality	12
1.3 Results	12
1.3.1 Search and Selection Procedures.....	12
1.3.2 Description of Studies	12
1.3.3 Measurement Properties and Methodological Quality.....	30
1.3.4 Characteristics of Questionnaires	31
1.3.4.1 Acceptance and Action Questionnaire (AAQ; Hayes et al., 2004).....	31

Table of Contents

1.3.4.2	Acceptance and Action Questionnaire - Revised (AAQ-II; Bond et al., 2011).....	32
1.3.4.3	Five Factor Mindfulness Questionnaire (FFMQ; Baer et al., 2006)	33
1.3.4.4	Five Factor Mindfulness Questionnaire-15 (Baer et al., 2012).....	36
1.3.4.5	Kentucky Inventory of Mindfulness Skills (Baer et al., 2004)	37
1.3.4.6	Philadelphia Mindfulness Scale (Cardaciotto et al., 2008).....	38
1.3.4.7	Difficulties in Emotion Regulation Scale (Gratz & Roemer, 2004).....	39
1.3.4.8	State Difficulties in Emotion Regulation Scale (Lavender et al., 2017).....	40
1.3.4.9	Affective Style Questionnaire (Hofmann & Kashdan, 2010).....	41
1.3.5	Summary of COSMIN methodological checklist	42
1.3.6	Summary of Terwee's Quality Criteria.....	43
1.4	Discussion.....	53
1.4.1	Summary of findings.....	53
1.4.2	Theoretical definitions of acceptance	53
1.4.3	Acceptance and Non-acceptance.....	55
1.4.4	Utility of Acceptance Scales	56
1.4.5	Use of COSMIN and Terwee Quality Criteria	56
1.4.6	Strengths.....	59
1.4.7	Limitations	59
1.4.8	Conclusion.....	60
 Chapter 2 Empirical Paper: Development of the Southampton Acceptance Scale		
	(SAS)	61
2.1	Introduction.....	61
2.1.1	Definitions of Acceptance	61
2.1.2	Acceptance and Commitment Therapy.....	64

2.1.3	Acceptance and ACT	65
2.1.4	Measuring acceptance	66
2.1.4.1	General versus specific measures of acceptance	67
2.1.4.2	Other acceptance measures.....	68
2.1.4.3	Process versus outcomes	69
2.1.4.4	Dimensionality of acceptance and experiential avoidance.....	70
2.1.4.5	Summary of limitations of current measures	71
2.1.5	Research Objectives.....	71
2.2	Phase One	72
2.2.1	Method.....	72
2.2.1.1	Preliminary item pool.....	72
2.2.2	Results and discussion.....	75
2.3	Phase Two.....	75
2.3.1	Method.....	78
2.3.1.1	Participants.....	78
2.3.1.2	Materials and procedure.....	78
2.3.2	Results and discussion.....	78
2.3.2.1	Factor Structure	78
2.3.2.2	Incentives	80
2.3.2.3	Reading Level	83
2.4	Phase Three.....	83
2.4.1	Method.....	85
2.4.1.1	Participants.....	85
2.4.1.2	Materials and procedure.....	86

Table of Contents

2.4.1.3	Analytic Strategy	89
2.4.2	Results and discussion.....	90
2.4.2.1	Confirmatory Factor Analyses	90
2.4.2.2	Validity of the SAS	95
2.4.2.3	Format invariance	96
2.5	General Discussion	97
2.5.1	Dimensionality.....	98
2.5.2	Reliability and validity	99
2.5.3	Strengths.....	101
2.5.4	Limitations	102
2.5.5	Directions for future research	102
2.5.6	Clinical implications and potential uses.....	104
2.5.7	Conclusions	104
Appendix A	COSMIN manual summary table for definitions of domains, measurement properties, and aspects of measurement properties.....	105
Appendix B	Quality criteria for measurement properties of health status questionnaires (Terwee et al., 2007).....	107
Appendix C	ACT Experts Information Sheet	110
Appendix D	Ethical Approval	112
Appendix E	Emotion Regulation and Wellbeing across the Lifespan: The Role of Life Stressors and Resilience Part One - Developing a new Questionnaire Information Sheet.....	113
Appendix F	Emotion Regulation and Wellbeing across the Lifespan: The Role of Life Stressors and Resilience Part One - Developing a new Questionnaire. Debriefing statement	115

Appendix G	Attitudes and Emotions Across the Lifespan: Information Sheet	117
Appendix H	Attitudes and Emotions Across the Life Span: Debriefing Statement	120
Appendix I	The Southampton Acceptance Scale (SAS)	123
Appendix J	Acceptance and Action Questionnaire- Revised (AAQ-II)	125
Appendix K	Emotion Regulation Questionnaire (ERQ)	126
Appendix L	Positive and Negative Affect Scale (PANAS)	127
Appendix M	Short Warwick Edinburgh Mental Well-being Scale (SWEMWBS).....	128
Appendix N	SWEMWBS Conversion Table	129
Appendix O	Balanced Inventory of Desired Responding – Short Form (BIDR-16), Impression Management Subscale	131
	List of References	133

Table of Tables

<i>Table 1.1</i>	<i>Search terms entered into databases</i>	<i>8</i>
<i>Table 1.2</i>	<i>Summary of Study Characteristics</i>	<i>15</i>
<i>Table 1.3</i>	<i>Description of Questionnaires and Subscales Measuring Acceptance .</i>	<i>27</i>
<i>Table 1.4</i>	<i>Methodological Quality of Each Sub-study per Measurement Property and Instrument.....</i>	<i>44</i>
<i>Table 1.5</i>	<i>Measurement Property Quality per Instrument.....</i>	<i>49</i>
<i>Table 2.1</i>	<i>Initial Exploratory Factor Analysis Factor Loadings Across a Non-Clinical Sample (n= 268)</i>	<i>76</i>
<i>Table 2.2</i>	<i>Factor loadings for Exploratory Factor Analysis with Direct Oblimin Rotation for Refined Item Pool.....</i>	<i>81</i>
<i>Table 2.3</i>	<i>Descriptive Statistics and Alpha Reliabilities Across Studies.....</i>	<i>82</i>
<i>Table 2.4</i>	<i>Demographics for Incentivised and Non-Incentivised Participants in Study Two</i>	<i>82</i>
<i>Table 2.5</i>	<i>Model comparisons for respecification of CFA model.....</i>	<i>92</i>
<i>Table 2.6</i>	<i>Goodness of Fit Statistics for Tests of Multigroup Invariance Across Age Groups.....</i>	<i>95</i>
<i>Table 2.7</i>	<i>Correlations Between the SAS Subscales With Existing Scales in Study Three</i>	<i>97</i>

Table of Figures

<i>Figure 1.1</i>	<i>Flow diagram of study selection based on PRISMA group guidance (Moher et al., 2009).....</i>	<i>13</i>
<i>Figure 2.1</i>	<i>Respecified model of factorial structure for the Southampton Acceptance Scale (SAS) Model 7.</i>	<i>93</i>

Academic Thesis: Declaration Of Authorship

I, Zoe M^cAndrews, declare that this thesis entitled ‘Why we need a better measure of acceptance: Development and initial validation of the Southampton Acceptance Scale’ and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signed:

Date:

Acknowledgements

I would like to take this opportunity to thank everyone who has helped to make this project possible. Firstly, I would like to thank all of the participants who generously gave their time to take part, and to the organisations who have shared and promoted this project.

Secondly, a special thank you goes to my research supervisors, Professor Lusía Stopa and Dr Claire Hart, whose support, advice and encouragement have been invaluable throughout every stage of this project. I would also like to thank my second rater, Jessica Richardson, for donating her time to support this project.

Finally, my heartfelt thanks go to my partner, Mike, to my family and friends, my mentor, and lastly, but by no means least, my wonderful cohort, who have all shown me unwavering support throughout this challenging and rewarding project.

Chapter 1 Psychometric Properties of Acceptance

Measures: A Systematic Review

1.1 Introduction

“Give us grace to accept with serenity the things that cannot be changed, courage to change the things that should be changed, and the wisdom to distinguish the one from the other” (Niebuhr, 1943, in Niebuhr & Brown, 1986, p.251). The Serenity Prayer is a well-known prayer by Reinhold Niebuhr which describes the tension between acceptance and change in life. The prayer appears in many forms throughout modern culture, and has been used by a number of organisations throughout history, including the armed forces and Alcoholics Anonymous.

The need to balance change with acceptance has more recently become a key focus in a number of psychological therapies. There is growing body of literature exploring the role of acceptance in promoting psychological wellbeing and reducing psychological distress. However, the robustness of the evidence base depends on good quality measurement tools and there is currently a wide range of questionnaires that are used to measure acceptance in the literature. This aim of this review is to evaluate whether these existing measures of acceptance are valid and reliable.

1.1.1 Emotion Regulation and Acceptance

Traditionally, models of psychopathology have identified failures in emotion regulation as one cause of difficulties in managing distress. Thus, many psychological therapies focus on teaching emotion regulation strategies and advocate changing experiences to alleviate distress. Third-wave Cognitive-behavioural therapies (CBT), such

Psychometric Properties of Acceptance Measures: A Systematic Review

as Acceptance and Commitment Therapy (ACT; Hayes, Strosahl & Wilson, 1999), however, suggest that dysfunctional attempts at emotion regulation can themselves lead to psychopathology.

ACT is grounded in Relational Frame Theory (RFT; see Hayes, Barnes-Holmes, & Roche, 2001), a behavioural theory which emphasises the role of language and cognition in human behaviour. RFT suggests that through language humans learn to respond relationally to stimuli, and emotional content becomes associated with these, such as objects, words or memories in a ‘relational frame’. As such, both the stimuli and thoughts about these stimuli trigger emotional reactions, which in turn elicit behavioural responses. When these emotional reactions are aversive or unpleasant, behavioural responses often aim to avoid such stimuli, or to control and reduce the associated distress. As more associations are made, the relational frames expand, leading to more stimuli being experienced as aversive and necessary to avoid. According to the theory, this expansion of aversive relationships has implications for both the development and maintenance of psychopathology.

In line with RFT, ACT posits that painful experiences are part of being human, and are not problematic in and of themselves. As such, they do not necessarily result in psychopathology. Instead, it is how one relates to such events that can result in, or exacerbate, psychopathology. ACT conceptualises some of these unhelpful responses as core ‘pathological processes’ and identifies six key processes within the ACT model. Each pathological process has a counter ‘therapeutic process’ and therapy seeks to cultivate the therapeutic processes so as to increase psychological flexibility (the ability to be open to experiences, be in the present moment and do what matters in line with values). These key processes are often represented using the ‘Hexaflex model’ (Harris, 2009).

‘Experiential avoidance’ is a key pathological process within the ACT model, which describes both difficulties in tolerating internal experiences, such as thoughts, emotions, memories, and the subsequent (often dysfunctional) attempts to avoid, regulate or control these in some way. Such attempts at avoidance often come with a behavioural cost, making it more difficult for individuals to engage in behaviours which move them towards their valued goals, known in ACT as committed action. This process of experiential avoidance and inaction is conceptualised as psychological inflexibility (Hayes, Levin, Plumb-Villardaga, Villatte, & Pistorello, 2013). Consequently, ACT identifies experiential avoidance as the source of difficulties, because it prevents an individual engaging in a meaningful life. Overcoming avoidance, rather than regulating emotions, is therefore the focus of ACT (Blackledge & Hayes, 2001).

Acceptance is the proposed antithesis of experiential avoidance; it is the active process of moving towards the internal experiences that arise with a willingness to embrace them. Acceptance advocates experiencing internal private events for what they are and without judgement. According to ACT, if an individual can do this without engaging in dysfunctional attempts to change, control or avoid events, they are more able to participate in meaningful activities. Acceptance does not preclude change however. ACT promotes balancing acceptance of private experiences and committed action in order to move towards a life worth living (change-oriented strategies; Hayes, 2004). Acceptance is therefore adopted not for its own sake, but to allow one to live a life that has value. Thus, while acceptance is not an emotion regulation strategy per se, it encourages an alternative way to experience and connect with emotions. Although not the aim of acceptance, the change in relationship to one’s experiences often results in changes in the experience or intensity of affect. As a result, the concept of acceptance has received considerable attention in the emotion regulation literature.

1.1.2 Acceptance across therapies

Although a key component of the ACT model, acceptance-based processes are not unique to ACT. Non-judgemental acceptance and present-moment awareness are well-established components of mindfulness, which has been taught for centuries. As psychological therapies have developed, many have emphasised the importance of acceptance.

Second-wave CBT, although conceptualised as a change-focussed therapy, promotes acceptance to support cognitive and behavioural change. For example, when employing exposure techniques, acceptance of anxiety symptoms is necessary to enable the individual to habituate to these and overcome avoidance. Furthermore, cognitive change strategies require that the individual first acknowledges and accepts the experience of distressing thoughts, before these can be worked on and challenged.

As CBT approaches have developed, mindfulness and acceptance have become a more central component of third –wave therapeutic models. One such therapy, Dialectical Behaviour Therapy (DBT; Linehan, 1993) promotes mindfulness and “radical acceptance”, and like ACT emphasises an acceptance-change dialectic. More recent cognitive therapy (CT) models, such as meta-cognitive therapy (Wells, 2009) and mindfulness-based cognitive therapy (MBCT; Segal, Williams & Teasdale, 2002) recommend observation and acceptance of mental events for what they are rather than trying to alter them. As such, symptom reduction is not the primary aim of third-wave approaches.

Although acceptance features in both second- and third-wave CBT approaches, there are important differences in how it is conceptualised, as well as in the proposed functions of acceptance within the models (Leahy, Tirch, & Napolitano, 2011). In second-wave CBT, the function of acceptance is to facilitate cognitive and behavioural change, with the aim of symptom reduction. ACT and DBT, on the other hand, promote acceptance to

enable an individual to live a full and meaningful life, whilst mindfully experiencing all of the emotions that come with that. Consequently, the theoretical orientation from which acceptance is approached influences responses to, and experiences of, affect.

Beyond psychological therapies, acceptance is a widely used term which is likely to hold multiple meanings in different population and contexts. As noted above, within ACT acceptance refers to an active process of openness and willingness towards experiences. However, this is not necessarily how the word is understood in common language outside this context. The Oxford dictionary defines acceptance as willingness to tolerate a difficult situation, giving the example sentence “*a mood of resigned acceptance*” (Acceptance [Def. 3], 2010). Furthermore, Harris (2009) notes that acceptance is often considered as synonymous with tolerance, resignation or passivity (“gritting your teeth and putting up with it” p.134). As such, he advocates that therapists avoid using the term acceptance when first starting therapy.

1.1.3 Measurement of Acceptance

Given these difficulties in defining acceptance, and communicating this definition, acceptance has proven a challenging process to operationalise and measure. As many measures of acceptance are developed within a particular theoretical framework, the way in which acceptance is defined and assessed varies across tools, and consequently may measure different latent constructs. Furthermore, the language used in self-report measures can give rise to multiple semantic interpretations beyond what was intended within a given context, resulting in further ambiguity as to what a particular tool is measuring. To support best practice in research and clinical settings, it is fundamental that the tools used to measure latent constructs, such as acceptance, are reliable and valid.

Psychometric theory is the process underlying psychological measurement, and provides a conceptual framework for evaluating the reliability, validity and utility of individual measurement instruments (Barker, Pistrang, & Elliott, 2016). Although there are several measures of acceptance, to date there has been no comprehensive assessment of their psychometric properties. The aim of the current systematic review is to evaluate existing measures of acceptance and generate evidence about which measures are the most appropriate for use in clinical and research settings when assessing acceptance.

We chose two established quality assessment tools to investigate the psychometric properties of a range of acceptance measures. Firstly, to measure methodological quality of reporting studies we used the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN; Mokkink et al., 2010a), which uses a taxonomy of measurement properties to systematically evaluate psychometric data of patient-reported outcome measures. COSMIN provides uniform standards and definitions to assess the methodological quality of studies under review. Quality criteria suggested by Terwee et al. (2007) were also used to evaluate the quality of psychometric properties of measurement instruments. This tool has been widely used in systematic reviews of psychometric properties to determine weaknesses in measurement properties. As low quality studies risk biasing the recommendation and selection of measurement tools, both COSMIN and Terwee et al.'s criteria were used to provide a comprehensive evaluation of study quality in addition to the strength of measurement properties being assessed.

1.1.4 Objectives

This systematic review sought to identify and critically evaluate psychometric properties of existing acceptance measures using standardised and established quality criteria. The identification of good quality tools will facilitate assessment and evaluation of

the processes of acceptance within both clinical and research settings. Furthermore, it is hoped that this review will highlight tools and measurement properties which require further evaluation and refinement, thus guiding future research.

1.2 Method

1.2.1 Search Strategy

Systematic searches of the following databases were made, up to 9th December 2017: EMBASE via OVID, PsycINFO via EBSCO, PsycARTICLES via EBSCO, CINAHL via EBSCO, MEDLINE via EBSCO, and Web of Science. Search terms were selected in relation to four key areas to find relevant studies examining measurement properties of self-report measures of acceptance (Table 1.1). Each of the four main filters were searched separately and then combined for the final search. The search strategy was developed based on guidance for performing systematic reviews of psychometric properties of measures (de Vet, Terwee, Mokkink & Knol, 2011).

The first author conducted the initial screening of titles and abstracts for all identified articles to determine those suitable for full text review. Full-text reviews were independently screened by two reviewers to ensure reliability. A plan for managing any discrepancies was defined a priori in line with best practice (Higgins & Green, 2011). Any disagreements were discussed by the two raters and resolved based on the eligibility criteria to reach a consensus. A third rater was available should a consensus not be reached. Inter-rater reliability analyses were performed using Cohen's kappa. Reference lists of selected studies were also hand-searched to identify other eligible studies.

Table 1.1

Search terms entered into databases

	Construct	Target Population	Instrument	Measurement Properties
Search terms	(Acceptance NR/ADJ15 Commitment) OR “experiential avoidance” OR “psychological flexibility” OR mindfulness	(Adult* OR elderly OR aged) (NOT child* NOT teen* NOT adolesc* NOT Animal)	Questionnaire OR Scale OR Measur* OR Assess* OR Self-report OR Inventory OR Instrument OR Index OR Checklist	Psychometr* OR Valid* OR Reliab* OR Develop* OR Respons* OR Interpret* OR Sensitiv* OR “Internal consistency” OR “Factor Analysis”

1.2.2 Eligibility Criteria

Studies were included if the aim was to develop or evaluate a self-report measure of acceptance in an adult population. Studies of measures which had acceptance subscales were included only if separate scores were determinable. Similarly, tools which measured avoidance or non-acceptance were only included if there was evidence that lower scores indicated acceptance and it was possible to calculate this. Initially, measures of psychological (in)flexibility were excluded; however, upon reviewing the identified articles, it was apparent that the Acceptance and Action Questionnaire- Revised (AAQ-II; Bond et al., 2011) was often used to evaluate acceptance and as such it was agreed to include it. To ensure sufficient evidence of measurement properties, only studies which evaluated two or more aspects of psychometric properties were included.

Studies were excluded if it was not possible to determine a separate score for acceptance, or if the evidence suggested that subscales should not be used independently. Tools which did not measure acceptance as a construct per se, but instead evaluated acceptance of a particular difficulty (such as chronic pain, eating, etc.) were excluded. Studies which assessed the structure of a construct without assessing the measurement properties of a particular tool were also excluded, as were studies of treatment effectiveness unless they included a specific aim to evaluate measurement properties. Studies were excluded if any of the sample was aged 16-years or less, if the instruments were used in any language other than English, or if the study was validating a translated version of the measure. Furthermore, studies that were published only as ‘grey literature’ were not included in this review.

1.2.3 Data extraction

Data extraction included key study characteristics, including country, sample size, participant demographics and descriptive statistics (age, gender, clinical or non-clinical, etc.) and questionnaire characteristics, such as measures or subscales used, item number and subscales and scoring method. Relevant measurement properties were extracted in line with guidance from the COSMIN checklist research group (Mokkink et al., 2010b) which included internal consistency, reliability (both test-retest and inter-rater), measurement error, content validity, structural validity, hypotheses testing, criterion validity and responsiveness.

1.2.4 Methodological Quality

The COSMIN checklist for systematic reviews was used to evaluate the methodological quality of studies (Terwee et al., 2012). Two raters independently assessed all included studies using the checklist to assess reliability. The COSMIN checklist is an

Psychometric Properties of Acceptance Measures: A Systematic Review

established standardised tool that has been validated for assessing the methodological quality of studies of measurement properties. The tool consists of nine domains each assessing a different measurement property: internal consistency, reliability (test-retest and inter-rater), measurement error, content validity, construct validity (separated into structural validity, hypothesis testing and cross-cultural validity), criterion validity, and responsiveness. Further information including definitions of each measurement property can be found in Appendix A. These measurement properties are grouped into three domains: reliability, validity and responsiveness.

Reliability concerns the degree to which repeated use of the measure produces similar scores across different conditions. This domain is made up of internal consistency (the degree of interrelatedness of individual items in a scale), test-retest reliability (consistency of scores within a participant considered to be stable across time), interrater reliability (degree of agreement in scoring by different raters) and measurement error (the degree to which changes in score are not the result of true change in the construct being measured).

Validity pertains to how well a tool measures the construct that it is attempting to measure. The validity domain consists of content validity, construct validity, cross-cultural validity and criterion validity. Construct validity is concerned with whether the measure accurately reflects the construct under study, and includes face validity, which is the degree to which an instrument looks as though it accurately reflects the construct being measured. Construct validity is separated into structural validity (how well do the scores accurately reflect the dimensionality of the construct being measured), hypothesis testing (the degree to which the measure performs as expected in relation to itself, other instruments, or different groups) and cross-cultural validity (how well translated or

culturally adapted measures adequately reflect the original tool). Criterion validity is the degree to which scores adequately reflect those of a gold standard measure.

Responsiveness describes the ability of a measurement tool to accurately detect change in the construct under assessment over time. Responsiveness may be assessed following clinical interventions to determine whether any meaningful change has occurred in the target domains.

In addition to the three domains of measurement properties, the COSMIN checklist also includes a further domain: interpretability. Interpretability is the degree to which qualitative meaning can be assigned to an instrument's quantitative scores or any change in these scores. This may include clinical cut-offs, severity ranges and clinical or reliable change in scores. While not a measurement property, interpretability of scores on an instrument is an important characteristic that is widely used, and so is rated separately on the COSMIN checklist.

When completing the COSMIN checklist, only boxes relating to a measurement property that has been considered are completed (i.e. it is not necessary to complete every box for every study); thus, the number of quality ratings available for each study will vary. Each section contains between 5 and 18 items, with 114-items in total. Items are rated using a 4-point Likert scale (excellent, good, fair or poor). The overall quality rating for each measurement property is defined using a "worst score counts" method (i.e. the lowest score in that section; Terwee et al., 2012).

Where multiple sub-studies were reported within a single paper, each sub-study was evaluated and rated separately for the given measurement properties. Furthermore, where studies assessed multiple measures simultaneously, each measure was independently evaluated and rated.

1.2.5 Measurement Property Quality

In addition to assessing methodological quality with COSMIN, each measurement property was assessed using the criteria for good measurement properties (Terwee et al., 2007; Appendix B). Each property is rated as either positive, indeterminate, negative or no information available. Again, articles with multiple studies or measures were evaluated separately and independently rated for addressed measurement properties.

1.3 Results

1.3.1 Search and Selection Procedures

The selection process is summarised in Figure 1.1. Initial database searches produced 5196 papers (3097 after removal of duplicates). Screening of titles and abstracts identified 63 potential studies for which full-texts were obtained and these were independently screened by two raters, after which a further 50 papers were excluded. Initial inter-rater reliability was substantial ($\kappa = .73$, $p < .001$; 90.48%). Disagreements were easily resolved and mutual consensus was reached for all studies. An additional seven papers were found through hand-searching reference lists of selected papers. Twenty eligible studies were identified in total, which are summarised in Table 1.2.

1.3.2 Description of Studies

Half of the studies included multiple samples, and the majority of samples had a greater proportion of females. Almost half the samples were student populations ($n=23$). Thirteen clinical samples across inpatient and outpatient settings were reported with a range of clinical presentations: depression ($n=3$), anxiety ($n=1$), mixed anxiety and depression ($n=1$), borderline personality disorder (BPD; $n=2$), substance misuse ($n=1$),

eating disorders ($n=1$), and not specified ($n=13$). Further characteristics of studies can be found in Table 1.2.

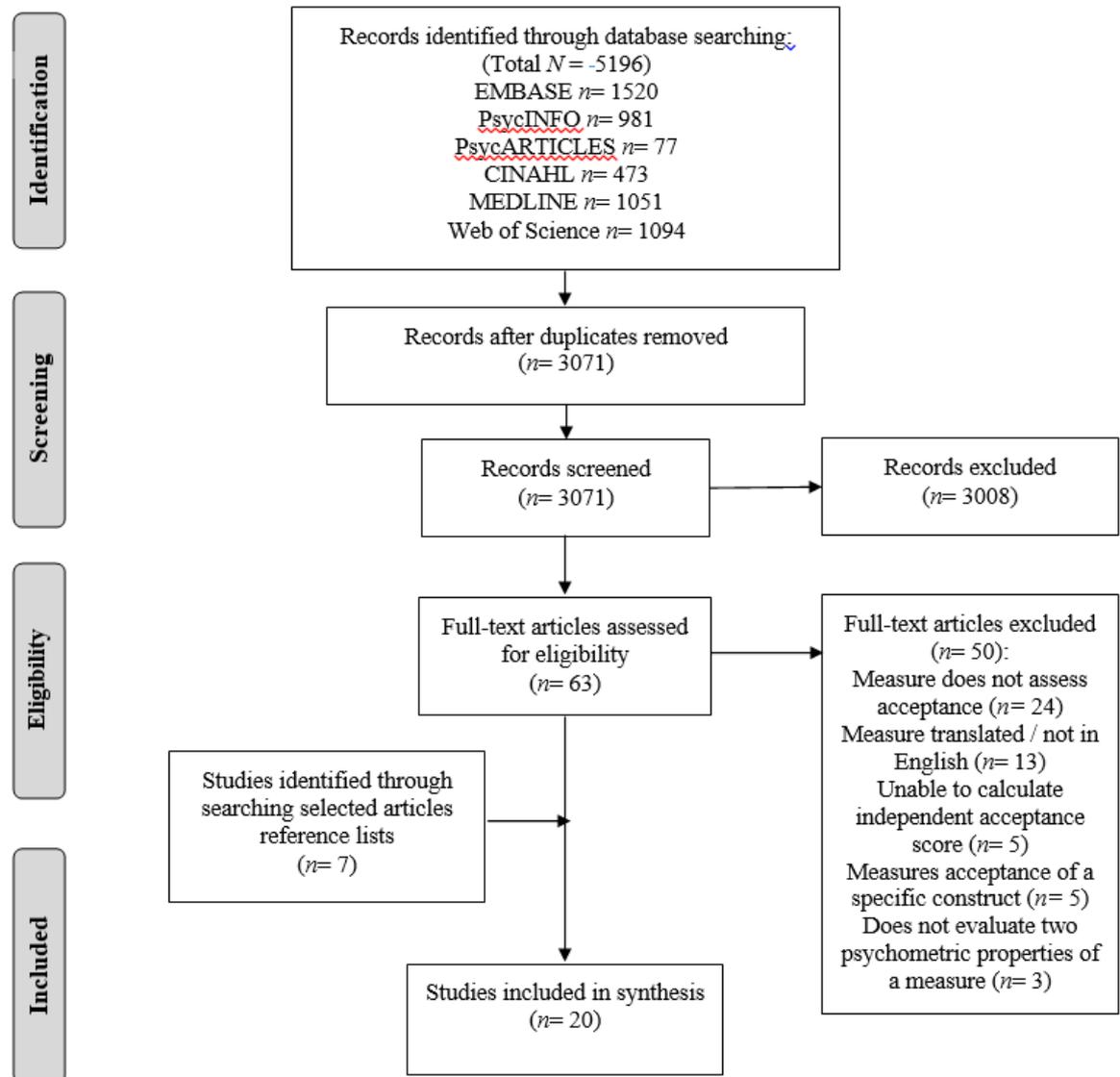


Figure 1.1 Flow diagram of study selection based on PRISMA group guidance (Moher et al., 2009).

Nine different measures were evaluated: four mindfulness measures, three emotion regulation measures and two measures of psychological flexibility/acceptance. The four

identified mindfulness scales were the Five Facet Mindfulness Questionnaire (FFMQ; Baer et al., 2006), the short form of the FFMQ (FFMQ-15; Baer, Carmody, & Hunsinger, 2012), the Kentucky Inventory of Mindfulness Skills (KIMS; Baer, Smith, & Allen, 2004) and the Philadelphia Mindfulness Scale (PHLMS; Cardaciotto, Herbert, Forman, Moitra, & Farrow, 2008). Measures of emotion (dys)regulation were the Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004), the state-version of the DERS (S-DERS; Lavender, Tull, DiLillo, Messman-Moore, & Gratz, 2017) and the Affective Style Questionnaire (ASQ; Hofmann & Kashdan, 2010). The psychological flexibility and acceptance scales were the Acceptance and Action Questionnaire (Hayes et al., 2004) and the revised version of the AAQ (AAQ-II; Bond et al., 2011). Before the AAQ was refined into the AAQ-II, several versions of this tool with varying item numbers were in use across research and clinical settings. Consequently, in the present review, several studies made use of different versions of the AAQ and this is referenced where it was reported in the original article. The characteristics of each measure are summarised in Table 1.3. No studies reported the time taken to complete the questionnaire, although all were easy to administer and calculate scores (total and subscales where appropriate).

Table 1.2

Summary of Study Characteristics

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	<i>N</i>	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Andrei et al. (2016)							FFMQ, PHLMS	IC, CVH
Sample 1	397	University Students	21.9(4.9)	305 Female	Caucasian (55.2%) Chinese (28.5%)	UK		
Sample 2	176	General population	36.7(14.4)	140 Female	Caucasian (82%)	NR		
Baer et al. (2004)								
Substudy 1	11	Experts	NR	8 Female	NR	NR	KIMS	CV
Substudy 2								
Sample 1	205	University Students	18-22	Female (60%)	Caucasian (85%)	USA	KIMS	IC
Sample 2	215	University Students	NR	NR	NR	USA	KIMS	
Sample 3	26	Adults with BPD	36, 20-52	25 Female	Caucasian	USA	KIMS	

Psychometric Properties of Acceptance Measures: A Systematic Review

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	<i>N</i>	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Substudy 3	49 (subsample of substudy 2 sample 2)	University Students	NR	NR	NR	USA	KIMS	TR
Substudy 4	130 (subsample of Study 2 sample 1)	University Students	19.6	Female (56%)	White (86%), Black (9%), Other (5%)	USA	KIMS	CVH, CVS
Substudy 5	115	University Students	NR	NR	NR	USA	KIMS	CVH
Substudy 6	Samples 1, 2 and 3 from substudy 2					USA	KIMS	CVH, CVS
Baer et al (2008)								
Sample 1	259	Students	18.9(3.2), 18-53	Male (22%)	White (92%)	USA	FFMQ	IC, CVH, CVS
Sample 2	293	Community	49.5(6.7), 34-66	Male (40%)	NR	UK		
Sample 3	252	Highly educated	44.2(11.9), 22-71	Male (42%)	White (91%)	USA		

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	<i>N</i>	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Baer et al (2008)								
Sample 4	213	Meditators	48.8(12.9), 18-83	Male (32%)	White (94%)	USA (80%)		
Bond et al. (2011)								
Substudy 1								
Sample 1	12	ACT experts	NR	NR	NR	Australia, Europe & USA	AAQ-II	IC, CV, CVS
Sample 2	26	PG students	NR	NR	NR	UK		
Sample 3	18	Community	NR	NR	NR	UK		
Sample 4	206	University Students	19(3.57)	Female (65%)	White/Caucasian (67%)	USA		

Psychometric Properties of Acceptance Measures: A Systematic Review

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	<i>N</i>	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Bond et al. (2011)								
Substudy 2							AAQ-II	IC, CVS
Sample 2	433	University Students	21(3.54)	Female (68%)	White (90%)	USA		
Sample 3	290	Clinical population – substance misuse	39(10.20)	Female (43%)	Caucasian (37%), Hispanic (29%), Black (28%)	USA		
Sample 4	583	Non-clinical community	34(9.76)	Female (58%)	White (97%)	UK		
Substudy 3							AAQ-II	TR, CVH
Samples 2-4 (above) plus								
Sample 5	872	Non-clinical University	35(9.47)	Female (58%)	White (95%)	UK		
Sample 6	432	students	19(2.63)	Female (100%)	White (72%)	USA		

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	<i>N</i>	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Cardaciotto et al. (2008)								
Substudy 1	6	Experts	-	4 Female			PHLMS	CV
Substudy 2	204	University students	21.9(3.83), 19-47	106 F, 94 M, 4 not-stated	White/Caucasian/European (64.7%), Asian/Pacific Islander (18.6%), Black/African American/Caribbean American (10.3%), Multiracial (5%), Hispanic/Latino/Latina (1.0%), other (0.5%)	USA	PHLMS	IC, CVH, CVS
Substudy 3	559	University students	20.12(3.49), 17-53	283 F, 270 M, 6 not stated	White/Caucasian/European (64.4%), Asian/Pacific Islander (19.0%), Black/African American/Caribbean American (8.1%), Multiracial (5.4%), Hispanic/Latino/Latina (1.6%), Native American (0.7%), other (0.7%)	USA	PHLMS	IC, CVH, CVS

Psychometric Properties of Acceptance Measures: A Systematic Review

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	<i>N</i>	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Cardaciotto et al. (2008)								
Substudy 4	52	Clinical outpatient population	40.78(12.04), 18-80	29 F, 23 M	Black/African American/Caribbean American (67.3%), White/Caucasian/European (19.2%), Hispanic/Latino/Latina (7.7%), Multiracial (5.8%),	USA	PHLMS	IC, CVH
Substudy 5	30	ED inpatient population (compared with sample from study 3)	30.0(10.60), 18-54	27 F, 3 M	White/Caucasian (90%)	NR	PHLMS	IC, CVH
Substudy 6	78	Students receiving psychotherapy	25.5(7.77), 18-49	69 F, 9 M	White/Caucasian (61.5%), Black/African American (11.6%), Multiracial/other (7.6%), Asian American (6.0%)	NR	PHLMS	IC, CVH

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	<i>N</i>	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Christopher et al. (2012)	349	Nonclinical sample	32.44(11.73)	263 F	White-non-Hispanic (80.2%), Black/African American (4.3%), Asian (3.7%), Hispanic/Latino (3.7%), Native America/Alaskan Native (1.1%), other (6.9%)	USA	FFMQ	IC, CVH, CVS
Curtiss and Klemanski (2014)	151	Clinical population (anxiety and depression)	28(14.8), 18-71	Female (63.6%)	White (79.5%), Hispanic (10.6%), African-American (6.6%), other (2.7%)	USA	FFMQ	IC, CVH
Goldberg et al. (2016)	130	Nonclinical sample	48.05(10.74)	79 F			FFMQ	IC, CV, CVH, Resp
Gratz and Roemer (2004)								
Substudy 1	357	University students	23.10(5.67), 18-55	Female (73%)	White (75%), Asian (17%), Black/African American (8%), Hispanic (4%)	USA	DERS	IC, CV, CVH, CVS

Psychometric Properties of Acceptance Measures: A Systematic Review

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	<i>N</i>	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Gratz and Roemer (2004)								
Substudy 2	21	University students	25.95(8.94), 18-48	Female (62%)	White (67%), Black/African American (24%), Asian/Pacific Islander (5%), unspecified (5%)	USA	DERS	TR
Gu et al. (2016)	238	Clinical sample – recurrent depression	49.18(12.01), 23-78	M=69 F= 169	White (97.5%)		FFMQ FFMQ-15	IC, CVS, CVH, Resp, CrV (FFMQ-15 only)
Hawley et al. (2017)	820	Clinical sample – anxiety disorders	33.56(10.85), 17-64	Female (57%)	Caucasian (80%)		FFMQ	IC, CVS
Hayes et al (2004)							AAQ	IC, CV, TR, CVS, CVH
Sample 1	460	Clinical sample	26	Female (63%)	Caucasian (85%)	NR		
Sample 2	419	Clinical sample	38.5	Female (65.4%)	NR	NR		

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	<i>N</i>	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Hayes et al. (2004)								
Sample 3	202	University students	20.5	Female (69.0%)	Caucasian (70.9%)	NR		
Sample 4	304	University students	19.0	Female (100%)	Caucasian (82.0%)	NR		
Sample 5	41	Clinical population – agoraphobia	38.0	Female (70.7%)	Caucasian (100%)	NR		
Sample 6	257	University Students	20	Female (82.0%)	Caucasian (82%)	NR		
Sample 7	205	University Students	22.5	Female (60.5%)	NR	NR		
Sample 8	51	Clinical sample – BPD inpatient	37.0	Female (100%)	Caucasian (80.0%)	UK		
Sample 9	97	Nonclinical – civil servants	21-36 (mean range)	Female (37.0%)	Caucasian (64.9%)	UK		
Sample 10	381	University students	18.9	F=179	NR	NR		

Psychometric Properties of Acceptance Measures: A Systematic Review

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	<i>N</i>	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Hofmann and Kashdan (2010)								
Substudy 1	434	University students	19.15(2.61)	Female (67%)	Caucasian (68.1%), Asian/Asian-American (18.4%), Hispanic/Latino/Mexican-American (3.7%), African American (2.6%)	USA	ASQ	IC, CV, CVS, CVH
Substudy 2	495	University students	22.02(5.23)	Female (78.0%)	Caucasian (54.5%), Asian/Asian-American (17.8%), Hispanic/Latino/Mexican-American (7.4%), African American (8.4%)	USA	ASQ	IC, CVS, CVH
Kortte et al. (2009)	139	Clinical population – physical injury/illness	54.90 (18.72), 18-92	Male (60.4%), Female (39.6%)	White (53.2%), African American (44.9%), Hispanic (1.4%), Asian/Pacific Islander (1.4%)	USA	AAQ	IC, CVS, CVH

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	<i>N</i>	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Lavender et al. (2017)	484	Community sample	21.8(2.2), <i>18-25</i>	Female (100%)	White (55.6%), African American/Black (32.2%), Latina (5.8%), Asian/Pacific Islander (2.7%)	USA	S-DERS	IC, CVS, CVH
Palladino et al. (2013)	238	Medical students	NR	NR	NR	USA	AAQ-II	IC, CVS, CVH
Scott et al. (2016)	573	Clinical sample – chronic pain	46.73 (11.29)	F=380, M=193	White (71.2%), Black (16.4%), Asian (7.3%), Mixed (4.2%)	UK	AAQ-II	IC, CVS
Siegling and Petrides (2016)							FFMQ PHLMS	IC, CVS, CVH
Sample 1	395	University students/ affiliates	21.9(4.9), <i>18-57</i>	Female (76.7%)	White (55.4%), East Asian (28.4%), south Asian (8.8%), other (8.4%)	UK		
Sample 2	172	Non-clinical sample	36.9(14.1), <i>18-76</i>	Female (79.7%)	Caucasian (84.3%), East Asian (2.9%), South Asian (1.7%), Black (4.7%), other/mixed (6.4%)	NR		

Psychometric Properties of Acceptance Measures: A Systematic Review

Authors	Population Characteristics					Country	Measure of Acceptance	Measurement Constructs Assessed
	N	Study Population	Age Mean Years, (SD), <i>range</i>	Gender	Ethnicity			
Watson-Singleton et al. (2018)	283	African-American clinical sample – suicide survivor	37.24 (12.24), 18-62	Female (52%), Male (42%), Not-stated (6%)	African American (100%)	USA	FFMQ	IC, TR, CVH
Williams et al. (2014)							FFMQ	IC, CVS
Sample 1	940	Non-clinical sample	25.7(9.8)	F=697 (74.1%)	White/Caucasian (85.1%), other (14.9%),	NR		
Sample 2	235	Meditators	46.51(13.1)	F=153 (65.1%)	White/Caucasian (91.9%), other (8.1%),	NR		
Sample 3	424	Clinical sample – recurrent depression	50.16(11.8)	F=325 (76.6%)	White/Caucasian (96.7%), other (0.9%), Not stated (2.4%)	UK		

Note. NR= not reported; AAQ = Acceptance and Action Questionnaire; AAQ-II = Acceptance and Action Questionnaire – Revised; FFMQ = Five Facet Mindfulness Questionnaire; FFMQ-15 = Five Facet Mindfulness Questionnaire-Short Form; KIMS= Kentucky Inventory of Mindfulness Skills; PHLMS= Philadelphia Mindfulness Scale; DERS= Difficulties in Emotion Regulation Scale; S-DERS= State Difficulties in Emotion Regulation Scale; ASQ= Affective Style Questionnaire; IC= Internal consistency; CV= Content Validity; CVH= Construct Validity – Hypothesis Testing; CVS= Construct Validity – Structural; TR= Test-Retest Reliability; Resp= Responsiveness; CrV= Criterion Validity.

Table 1.3

Description of Questionnaires and Subscales Measuring Acceptance

Measure	Construct Measured	Structure	Number of items*	Response options	Range of scores*	Scoring method*
AAQ	Experiential Avoidance	Single scale	9	7-point Likert scale (1= never true, 7= always true)	9-63	Items summed
AAQ-II	Psychological Inflexibility	Single scale	7	7-point Likert scale (1= never true, 7= always true)	7-49	Items summed
FFMQ	Mindfulness	Total score and 5 subscales: Non-judging Observing Describing Acting with Awareness Non-reactivity	8	5-point Likert scale (1=rarely or very rarely true, 5= very often or almost always true)	8-40	Items all reverse scored and summed
FFMQ-15	Mindfulness	Total score and 5 subscales: Non-judging Observing Describing Acting with Awareness Non-reactivity	3	5-point Likert scale (1=rarely or very rarely true, 5= very often or almost always true)	3-15	Items all reverse scored and summed
PHLMS	Mindfulness	2 subscales: Present Moment Awareness	10	5-point Likert scale (1=never, 5= very often)	10-50	Items all reverse scored and summed

Non-judgemental Acceptance

Measure	Construct Measured	Structure	Number of items*	Response options	Range of scores*	Scoring method*
KIMS	Mindfulness	Total score and 4 Subscales: Accept with Nonjudgement Observe Describe Act with awareness	9	5-point Likert scale (1=never or very rarely true, 5= always or almost always true)	9-45	Items all reverse scored and summed
DERS	Difficulties in Emotion Regulation	Total score and 6 Subscales: Non-acceptance Goals Impulse Awareness Strategies Clarity	6	5-point Likert scale (1=almost never, 5=almost always)	6-30	Items summed
S-DERS	Difficulties in Emotion Regulation	Total scale and 4 subscales: Non-acceptance Modulate Awareness Clarity	7	5-point Likert scale (1=not at all, 5=completely)	7-35	Items summed
ASQ	Emotion Regulation	3 Subscales: Tolerating Concealing Adjusting	5	5-point Likert scale (1=not true of me at all, 5=extremely true of me)	5-25	Items summed

Note: subscales in bold indicate relevant subscales evaluated in the present study; *for multidimensional measures information relates to relevant subscale, not the total scale; AAQ = Acceptance and Action Questionnaire; AAQ-II = Acceptance and Action Questionnaire – Revised; FFMQ = Five Facet Mindfulness Questionnaire; FFMQ-15 = Five Facet Mindfulness Questionnaire-Short Form; KIMS= Kentucky Inventory of Mindfulness Skills; PHLMS= Philadelphia Mindfulness Scale; DERS= Difficulties in Emotion Regulation Scale; S-DERS= State Difficulties in Emotion Regulation Scale; ASQ= Affective Style Questionnaire.

1.3.3 Measurement Properties and Methodological Quality

Inter-rater agreement of the COSMIN methodological quality ratings was almost perfect (ICC= .963 [.944, .975], $p < .001$), as were assessments of measurement property quality using Terwee et al.'s criteria (2007; ICC= .984 [.980, .988], $p < .001$).

Most studies assessed internal consistency and construct validity, whereas reliability and content validity were less commonly evaluated. As this review investigated measures in English only, no translated or culturally adapted versions of measures were included; it was therefore not possible to assess cross-cultural validity. Since there is no 'gold standard' for acceptance measures, it was not possible to assess criterion validity for the majority of studies. However, the COSMIN taxonomy states when validating a shortened version of an instrument, the longer version can be used as a 'gold standard' for comparison. Thus, criterion validity was assessed for one measure in this context. None of the studies assessed measurement error. Table 1.2 summarises the different measurement properties considered in each study. Methodological quality ratings based on the COSMIN taxonomy are summarised in Table 1.4. Many studies failed to report how missing data was handled, which resulted in lower scores across domains.

Given that the COSMIN rating employs a "worst score counts" methodology, there was a risk that otherwise high-quality research would be rated down for a single low score, resulting in such studies being conflated with those of overall lower quality. To better understand the range of scorings within sub-studies the authors evaluated the percentage of excellent ratings that were awarded across measurement properties. These results are summarised in Table 1.4.

Strength of psychometric properties was assessed using criteria suggested by Terwee et al. (2007), and is summarised in Table 1.5. All studies reported internal consistency,

whereas no studies reported floor and ceiling effects or agreement. No studies evaluated criterion validity in line with Terwee et al.'s criteria. Indeterminate ratings were often given due to a lack of target population involvement (content validity), absence of, or suboptimal factor analyses (internal consistency), lack of clear hypotheses (construct validity), results for fewer than four relevant subgroups of patients (interpretability), and inappropriate time intervals or design flaws (reliability). Results for each measurement tool are summarised below.

1.3.4 Characteristics of Questionnaires

1.3.4.1 Acceptance and Action Questionnaire (AAQ; Hayes et al., 2004)

The AAQ was the first tool designed to measure acceptance and experiential avoidance, as there was no suitable measure for emerging behaviour therapies which emphasised these constructs. Lower scores on the AAQ indicate greater acceptance and less avoidance (Hayes et al., 2004). Methodological quality across the two studies assessing the AAQ was fair-to-poor, with lack of clarity and detail in reporting resulting in lower ratings (Hayes et al., 2004; Kortte, Veiel, Batten, & Wegener, 2009). The AAQ exists in several forms (16-, 9-, and 7-item versions), all of which were evaluated in the development study (Hayes et al., 2004), while Kortte et al. (2009) used the 9-item version. Due to lack of clarity in the reports of Hayes et al. regarding which version was used in each analyses across samples, the findings of this study are reported together.

Hayes et al. (2004) recommended a single factor structure for the 9-item version of the AAQ; however, Kortte et al. (2009) reported that a two-factor solution was more suited in medical rehabilitation populations (although internal consistency for each factor was not reported). Internal consistency for the scale was acceptable ($\alpha=.70$ in both studies). Moderate correlations were found with thought suppression ($r= .44$ and $.50$) and thought

control scales ($r = .36$ and $.37$), supporting construct validity of the AAQ (Hayes et al., 2004). However, much data used to evaluate the AAQ's relationship to other measures came from other studies. As such, the AAQ is unlikely to have been standardised across studies which limits comparisons that can be drawn. For example, sample seven were administered a 16-item "experiential avoidance scale", whereas sample two completed an "acceptance index", and sample nine completed the 16-item AAQ.

Neither study evaluated the responsiveness of the AAQ. Test-retest reliability was assessed for the 16-item version over a 4-month period. However, the report does not state which method was used to evaluate test-retest reliability (intraclass correlation coefficient (ICC) or Pearson correlation). Furthermore, no rationale is given for selecting this time period, which is longer than is often used with self-report measures of this nature. As it is not possible to determine whether the participants were stable over a timeframe which was longer than expected, test-retest reliability is rated as methodologically poor.

1.3.4.2 Acceptance and Action Questionnaire - Revised (AAQ-II; Bond et al., 2011)

The AAQ-II was developed in response to criticisms of the AAQ's poor internal consistency and construct validity (Bond et al., 2011). The AAQ-II was designed to be a broader measure of psychological (in)flexibility, that could be used across a wide range of settings. As such, the AAQ-II aims to tap into various aspects of the hexaflex model beyond just acceptance and experiential avoidance. However, as the original AAQ was developed as an acceptance/experiential avoidance measure, the AAQ-II has continued to be used in the same way. Higher scores indicate greater inflexibility/avoidance, whereas lower scores represent greater flexibility/acceptance.

Methodological quality of studies evaluating the AAQ-II ranged from excellent to poor (Hayes et al., 2004; Palladino et al., 2013; Scott, McCracken, & Norton, 2016).

However, lower ratings were often the result of ambiguity or lack of detail in reporting (such as not specifying the statistical analyses used to evaluate reliability or vague hypotheses). The AAQ-II is a unidimensional measure, and this factor structure was confirmed by both of the studies that evaluated it. Internal consistency was acceptable to excellent across all articles (Cronbach's alpha ranging from .78-.93). Construct validity was supported; the AAQ-II demonstrated a strong negative correlation with the Mindfulness Attention and Awareness Scale (MAAS; $r=-.53$), and strong positive correlations with the Cognitive Fusion Questionnaire (CFQ; $r=.63$; Palladino et al., 2013). In support of concurrent validity, there were strong positive correlations for suppression ($r=.63$), stress ($r=.57$) and depression ($r=.61$; Bond et al., 2011). By comparison, there were moderate positive correlations with anxiety ($r=.49$), and general mental health difficulties (r ranged from .30-.53; Bond et al., 2011). There was no relationship between scores on the AAQ-II and age, gender or ethnicity (Bond et al., 2011). None of the studies evaluated the responsiveness of the AAQ-II. Test-retest reliability was assessed in a non-clinical population of UK retail bank workers across 3- and 12-months. Although authors report substantial reliability of .81 and .79 respectively, it is not clear whether ICC or Pearson's correlations were conducted. ICC is the preferred analysis identified in the quality criteria as it considers systematic error, unlike correlations. Furthermore, as it is unclear whether participants were stable across this time period, this was rated as methodologically fair.

1.3.4.3 Five Factor Mindfulness Questionnaire (FFMQ; Baer et al., 2006)

The FFMQ was developed following evaluation of the facets of mindfulness whereby a factor analysis was conducted on the combined item pool of five established measures of mindfulness (including the KIMS; Baer et al., 2004). The FFMQ was the most widely evaluated measure in this review, with ten studies assessing psychometric

properties across a range of clinical (depression, anxiety and suicide survivors) and non-clinical (students, community and meditator) samples. Studies varied considerably in methodological quality across the measurement properties (from excellent to poor). Many studies investigated the factor structure of the FFMQ. In the development paper, Baer et al. (2006) recommended a five-factor structure assessing the key constructs of mindfulness. However, other research supports a four-factor solution due to poor internal consistency and model fit in confirmatory factor analyses (CFA; e.g. Siegling & Petrides, 2016). Williams, Dalgleish, Karl, & Kuyken (2014) argued that the optimal factor structure varied depending on meditation experience, with a four-factor solution being preferred in samples without meditation experience. This finding was replicated by Gu et al. (2016), who compared the factor-structure pre- and post-MBCT, with four and five-factor models providing better fits respectively. Notwithstanding discrepancies in model fit, the non-judging factor has been retained across all analyses.

The non-judging scale measures acceptance of experience, containing items relating to judgements and self-criticism, all of which are reverse scored and many of which originate from the KIMS. Internal consistency was in the recommended range ($\alpha=.70$ to $.95$) for the majority of studies and sample populations. The exceptions to this were the African-American suicide survivor sample in Watson-Singleton, Walker, LoParo, Mack, and Kaslow (2018; $\alpha=.69$), and the study of Andrei, Vesely, and Siegling (2016) who did not perform internal consistency analyses for individual facets.

Construct validity was assessed in a number of studies, which yielded mixed results. The non-judging subscale had strong positive correlations with the PHLMS acceptance scale in multiple studies ($r=.58$ and $.72$, Siegling & Petrides, 2016; $r=.58$ and $.72$, Andrei et al., 2016). Non-judging was positively associated with a self-compassion measure ($r=.31$), and negatively correlated with self-criticism ($r=-.42$) and self-judgement ($r=-.56$;

Watson–Singleton et al., 2018). Curtiss and Klemanski (2014) found that non-judging was strongly associated with the AAQ-II ($r=.54$). Higher scores on the AAQ-II indicate greater psychological inflexibility, and as such, a negative correlation would be expected with non-judging. It is possible that the authors reverse scored the AAQ-II so that higher scores indicated greater flexibility, which would support the conclusions drawn; however, no information is provided in the paper regarding this process. Furthermore, the non-judging scale was not significantly correlated with the DERS nonacceptance scale, or the ERQ suppression scale. Curtiss and Klemanski's study was rated as 'good' (COSMIN) and + (positive; Terwee et al., 2007) for hypothesis testing and construct validity respectively. This was the only study to employ false discovery rate (FDR) procedures to compensate for the large number of correlation analyses. FDRs, which differ from family-wise error processes such as Bonferroni corrections, are the proportion of null hypotheses that are falsely rejected, aiming to control for this whilst retaining power (Benjamini & Yekutieli, 2001). Taken together, these results suggest that further investigation of convergent validity in clinical populations is warranted.

Two studies compared FFMQ scores across multiple groups. Hawley et al. (2017) found significant differences in non-judgement scores across anxiety diagnostic groups. Individuals with a primary diagnosis of panic disorder (with or without agoraphobia) scored significantly higher compared with those diagnosed with generalised anxiety disorder (GAD; a disorder characterised by worry and intolerance of uncertainty). Baer et al. (2008) compared scores of meditators and non-meditators, to determine whether the FFMQ was sensitive to the cultivation of mindfulness skills practised in meditation. As expected, scores for non-judgement were significantly higher in meditators compared to students and a demographically matched sample.

Responsiveness was assessed by two studies. Gu et al. (2016) found that non-judging scores significantly increased pre- and post-MBCT, with a large effect size. Goldberg et al. (2016) also compared scores pre- and post-interventions (Mindfulness-based stress reduction (MBSR), an active control and waiting-list control). All three groups showed significant increases in scores for non-judging across the two time-points. However, there was no significant difference in responsiveness between the MBSR and the active control group. Although these studies present conflicting evidence for the responsiveness subscale of the FFMQ, it should be noted that both compared pre- and post- scores using paired-sample t-tests and effect sizes, whereas Terwee et al. (2007) recommend either relating the smallest detectable change (SDC) to the minimal important change (MIC) in scores, or assessing the area under the receiver operating characteristics (ROC) curve (AUC). As a result, both studies were given an 'indeterminate' rating because it is possible that their analysis was not sensitive enough to detect measurement error.

1.3.4.4 Five Factor Mindfulness Questionnaire-15 (Baer et al., 2012)

The FFMQ-15 was developed to reduce participant burden whilst maintaining the breadth of the original measure in a study evaluating a mindfulness intervention (Baer et al., 2012). As such, the FFMQ-15 uses the same five subscales, although each has fewer items. Only one study evaluated the FFMQ-15, and it had fair methodological quality (Gu et al., 2016). However, this was as a result of failure to report handling of missing data, which may have negatively biased results (the study would otherwise have received good ratings). Like the original FFMQ, CFA pre- and post-MBCT showed discrepancies in the best model fit (with four and five factor solutions recommended respectively), which raises questions about the construct validity of the scale when used in populations who are not familiar with mindfulness. Internal consistency for the non-judging scale was acceptable at both time points ($\alpha=.76$ and $.78$). The FFMQ-15 was sensitive to change over time

measured by correlations. Criterion validity was evaluated using the COMSIN criteria as it was possible to use the FFMQ as the ‘gold standard’ (but not for Terwee et al.’s (2007) criteria). The non-judging subscales on both versions of the FFMQ were highly correlated ($r=.90$). Construct validity was supported for this subscale by moderate to large negative correlations with depression ($r=-.30$ and $-.42$) and rumination ($r=-.50$ and $-.58$).

1.3.4.5 Kentucky Inventory of Mindfulness Skills (Baer et al., 2004)

The KIMS was designed to assess mindfulness in everyday life for use with those unfamiliar with the concept or unpractised in meditation (Baer, Smith & Allen, 2004). Only one study in this review evaluated the KIMS and the methodological quality was fair-to-poor across measurement properties (Baer et al., 2004). The KIMS consists of 39-items across four factors which correspond with skills taught in DBT: ‘observing’, ‘describing’, ‘acting with awareness’ and ‘accepting without judgement’. Factor analyses confirmed the 4-factor solution. Pertinent to this review is the “acceptance without judgement subscale” (AWJ), which aims to measure an individual’s ability to accept the presence of unwanted experiences without trying to change or eliminate them. Items ask about making judgements and self-criticism, and all items are reverse scored. This subscale had good internal consistency ($\alpha=.87$). The construct validity of the AWJ was supported by a small negative correlation with the AAQ in students ($r=-.26$) and significant differences between AWJ scores for students and a sample of people with BPD were found (however, the BPD sample was underpowered, which may have implications for interpretation). In total, 19 scales are used to determine convergent and concurrent validity, the authors attempt to compensate for multiple correlations by changing the significance level to $p>.01$. However, no rationale is given for selecting this significance value. Test-retest reliability was demonstrated for AWJ (.83); however, the sample size was small and underpowered ($n=49$), and so this was rated as fair.

1.3.4.6 Philadelphia Mindfulness Scale (Cardaciotto et al., 2008)

The PHLMS was developed with the aim of assessing a bi-dimensional model of mindfulness in line with Kabat-Zinn's definition (1994): present moment awareness and non-judgemental acceptance. The subscales were uncorrelated, meaning a total score is not indicated, and the subscales should therefore be interpreted separately. The acceptance scale aims to measure a non-judgemental stance towards experiences (including emotions and thoughts). Three studies in this review evaluated the PHLMS and methodological quality varied from good to poor (Andrei et al., 2016; Cardaciotto et al., 2008; Siegling & Petrides, 2016).

Content validity was rated as poor in the development study, as only experts were involved in item generation and selection, and there was no assessment of the comprehensiveness of items (Cardaciotto et al., 2008). Despite a small sample size, the exploratory factor analyses (EFA) indicated a two-factor solution, and this was supported by CFA, for which the sample size was adequate (Cardaciotto et al., 2008). Andrei et al. was the only study that did not conduct factor analyses. Internal consistency for the non-judgemental acceptance scale was good across all studies (Cronbach's alphas range from .82 to .88).

Construct validity was supported in all three studies; Siegling and Petrides (2016) and Andrei et al. (2016) demonstrated strong correlations between the acceptance subscale and the AWJ subscale of the FFMQ ($r=.58$ and $r=.58-.72$ respectively). Siegling and Petrides also found that the non-judgemental acceptance subscale loaded onto the same factor as the AWJ subscale of the FFMQ, suggesting that they are measuring the same latent variable. Cardaciotto et al. (2008) assessed construct validity across three samples (non-clinical students, general clinical population and student counselling centre sample). All three samples demonstrated moderate-to-large correlations with the BDI ($r=-.35, -.28$

and -.51 respectively) and the BAI ($r=-.33$, $-.29$ and $-.39$ respectively). There were also strong correlations with the AAQ (which was reverse scored to measure acceptance in line with directionality of other measures in the analyses; $r=.54$) and the White Bear Suppression Inventory (Wegner & Zanakos, 1994; $r=-.52$) in the non-clinical sample. In an attempt to correct for type I errors across multiple correlations, the Cardaciotto et al. study considered relationships significant if $p<.01$. However, they failed to provide a rationale for choosing this significance level. Whilst Bonferroni corrections are often considered overly conservative (Perneger, 1998), given the small sample size in some of the populations assessed ($n= 52$), a more robust approach to type I errors would have been preferable. Notwithstanding, the pattern of correlations with other measures was consistent across all samples, suggesting further assessment of construct validity with larger samples would be of value. None of the studies evaluated the test-retest reliability or sensitivity of the PHLMS.

1.3.4.7 Difficulties in Emotion Regulation Scale (Gratz & Roemer, 2004)

The DERS was developed to assess a range of emotion regulation difficulties associated with the development and maintenance of psychopathology using a single measure to reduce patient burden and clinical overlap which can occur when using multiple tools (Gratz & Roemer, 2004). Acceptance of emotions was a key dimension of emotion regulation that the tool aimed to assess. The nonacceptance of emotional responses subscale measures how negative or nonaccepting individuals are of their emotional responses and distress, with higher scores indicating less acceptance. The only study to evaluate the psychometric properties of the DERS was the development study of Gratz and Roemer, with methodological quality of fair-to-poor across measurement properties.

Content validity was rated as poor (COSMIN) and indeterminate (Terwee) due to lack of detail about the item generation and selection process. The DERS six-factor solution was identified by EFA, but no CFA was reported, so the fit of this solution is yet to be tested. Internal consistency for the nonacceptance subscale was good ($\alpha=.85$). An independent sample was used to determine test-retest reliability; however, the sample was very small ($n=21$) and the time period varied across participants (from 4-8weeks). As such it is not possible to determine the reliability of this measure over time, resulting in poor and indeterminate ratings for COSMIN and Terwee criteria respectively. Construct validity was supported for the nonacceptance subscale, which showed a moderate correlation with experiential avoidance ($r=.39$). Three measures were used to assess construct validity, and a number of separate correlations were performed. Despite this, no attempts to compensate for family-wise error were made, suggesting that significant associations might be inflated, although the large sample size ($n=357$) might offer some protection from this.

1.3.4.8 State Difficulties in Emotion Regulation Scale (Lavender et al., 2017)

The S-DERS was developed to assess emotional regulation difficulties at a particular moment. The authors argued that many factors can influence emotion dysregulation within short time frames and most scales are not sensitive enough to detect this due to their focus on dispositional tendencies. The S-DERS was based on the DERS, with items modified to assess each domain in a momentary fashion. The only study to evaluate the psychometric properties of the S-DERS was the development study, which had a methodological quality of good-to-poor across measurement properties.

Despite being styled on the DERS, EFA demonstrated a different factor solution; four of the original six factors remained, one of which was the non-acceptance scale. Individuals with missing data were excluded from the analyses, and no evidence of missing data analysis is provided, which might result in bias. Furthermore, no CFA was

undertaken, as such this proposed structure is yet to be confirmed. However, internal consistency of the non-acceptance scale was excellent ($\alpha=.92$). Non-acceptance demonstrated convergent validity, showing significant moderate-to-large correlations with the DERS non-acceptance subscale ($r=.44$) and the AAQ ($r=.41$).

1.3.4.9 Affective Style Questionnaire (Hofmann & Kashdan, 2010)

The ASQ was developed to measure three broad styles of affect regulation that were not restricted to a particular theoretical orientation or to specific strategies. The development paper was the only study to evaluate the ASQ in the current review. Content validity was rated as methodologically ‘poor’, due to lack of detail regarding the item generation and evaluation processes. Authors stated items were largely informed by Gross and John’s work on the process model of emotion regulation (1997, 2003) and on the acceptance and mindfulness literature (Hayes et al., 1999), however, they did not expand upon this process, resulting in a rating of poor content validity. Principal component analysis (PCA) was used in place of factor analysis, which is not recommended in scale development as PCA is a data reduction method rather than seeking to understand relationships of items with latent variables (Worthington & Whittaker, 2006). While some researchers argue PCA and EFA often produce equivalent results, PCA is not a true measure of underlying structures, and as such might have implications for interpretation of the structure of the ASQ (see Russell (2002) for a review). Furthermore, an insufficient sample was used in this phase of analysis based on COSMIN and Terwee recommendations. The PCA identified three components, the third of which was labelled ‘tolerating’. The authors describe the tolerating scale as similar to acceptance, in that it reflects non-defensive responses to emotional experiences, including a strong tolerance of distress. The authors conducted PCA in two independent studies, both of which used demographically similar student samples. Consequently, although study two replicated the

findings of study one, a CFA would have been more appropriate to test the proposed three-facet structure of the ASQ. Internal consistency was questionable in both samples ($\alpha=.66$ and $.68$). However, tolerating is the smallest scale of the ASQ at only 5-items, which may partially explain poorer internal consistency (DeVellis, 2003).

Construct validity was demonstrated across both samples by significant moderate negative correlations with suppression on the ERQ ($r=-.34$ and $-.32$), and small-to-moderate correlations with the non-acceptance subscale of the DERS ($r=-.31$ and $-.15$). However, the tolerating subscale demonstrated significant positive correlations with the AAQ-II (10-item version; $r=.22$ and $.18$) suggesting some overlap with experiential avoidance rather than acceptance (as the authors do not suggest that they reverse scored the AAQ-II). Furthermore, a substantial number of correlations were made, and no evidence was provided regarding attempts to control for type I errors, which might have resulted in spurious significant associations. Consequently, construct validity methodology was rated as poor.

1.3.5 Summary of COSMIN methodological checklist

The COSMIN checklist enabled a robust and detailed evaluation of the methodological quality of studies using acceptance scales. Common limitations across studies were not reporting missing data handling, insufficient sample size, inappropriate time-frames between measurements and/or lack of detail about stability of participants in this period. Furthermore, lower ratings were also given due to ambiguity in reporting, such as what statistical methods had been used in analysis or which version of a scale was used.

The COSMIN tool uses a ‘worst score counts’ methodology for each area of assessment, and as such, most studies were rated as fair or poor. To explore whether there was potential negative bias on the ratings of studies due to the worst score counts

approach, the authors evaluated the percentage of excellent ratings achieved. These are summarised in Table 1.4. Scott et al. (2016) achieved the highest rating (92.31%) assessing the AAQ-II, however, this study only received ‘fair’ ratings in both measurement properties that were assessed. This was followed by Hawley et al. (2017) who received 86.96% of excellent ratings, which resulted in two ‘excellent’ and one ‘fair’ overall ratings across assessed properties for the FFMQ. The lowest rating was 29.41% for Andrei et al. (2016) who received two overall ‘poor’ ratings, also assessing the FFMQ. The considerable range in excellent scores suggests that while some studies may suffer from negative bias in the worst score counts approach, this is not a universal issue.

1.3.6 Summary of Terwee’s Quality Criteria

Terwee’s criteria enabled quality assessment of overall measurement properties, which are summarised in Table 1.5. Four studies received negative ratings for content validity, due to lack of target population involvement during item selection. Two studies received negative ratings for internal consistency, as Cronbach’s alpha was below the recommended .70: Watson-Singleton et al. (2017, $\alpha = .69$) and Hofmann and Kashdan (2010; $\alpha = .66$).

Studies were rated as ‘0’ when no information about that measurement property was provided. Indeterminate (“?”) ratings were given for concerns regarding the design or methodology of the study, lack of clear hypotheses (construct validity), not conducting a factor analysis (internal consistency), insufficient groups (interpretability) and use of non-recommended statistical analyses (reliability and responsiveness).

Table 1.4

Methodological Quality of Each Sub-study per Measurement Property and Instrument

Measure	Internal consistency	Reliability	Content validity	Construct validity		Criterion validity	Responsiveness	Excellent Ratings (%)
				Structural validity	Hypothesis testing			
AAQ								
Hayes et al (2004)	Fair	Poor	Poor	Fair	Poor			33.33
Kortte et al. (2009)	Fair			Fair	Fair			56.52
AAQ-II								
Bond et al. (2011)								
Study 1	Fair		Excellent	Fair				77.78
Study 2	Fair			Fair				69.23
Study 3		Fair			Poor			38.10
Palladino et al. (2013)	Excellent			Excellent	Poor			82.61
Scott et al. (2016)	Fair			Fair				92.31
FFMQ								
Andrei et al. (2016)	Poor				Poor			29.41

Measure	Internal consistency	Reliability	Content validity	Construct validity		Criterion validity	Responsiveness	<i>Excellent Ratings (%)</i>
				Structural validity	Hypothesis testing			
FFMQ								
Baer et al. (2008)	Good			Good	Fair			63.64
Christopher et al. (2012)	Fair			Fair	Poor			60.87
Curtiss and Klemanski (2014)	Poor			Poor	Good			73.91
Goldberg et al. (2016)	Fair				Fair		Fair	62.50
Gu et al. (2016)	Fair			Fair	Fair		Fair	60.87
Hawley et al. (2017)	Excellent			Excellent	Poor			86.96
Siegling and Petrides (2016)	Poor			Poor	Poor			52.63
Watson-Singleton et al. (2018)	Fair	Poor		Fair	Fair			64.71
Williams et al. (2014)	Fair			Fair				53.85
FFMQ-15								
Gu et al. (2016)	Fair			Fair	Fair	Fair	Fair	63.16

Measure	Internal consistency	Reliability	Content validity	Construct validity		Criterion validity	Responsiveness	Excellent Ratings (%)
				Structural validity	Hypothesis testing			
KIMS								
Baer et al. (2004)								
Study 1			Poor					80.00
Study 2	Fair			Fair				53.85
Study 3		Fair						54.55
Study 4					Fair			37.50
Study 5					Fair			40.00
Baer et al. (2004)								
Study 6					Poor			50.00
PHLMS								
Andrei et al. (2016)								
	Fair				Poor			41.18
Cardaciotto et al. (2008)								
Study 1			Poor					60.00
Study 2	Poor			Poor				50.00

Measure	Internal consistency	Reliability	Content validity	Construct validity		Criterion validity	Responsiveness	<i>Excellent Ratings (%)</i>
				Structural validity	Hypothesis testing			
PHLMS								
Cardaciotto et al. (2008)								
Study 3	Fair			Fair	Fair			65.22
Study 4	Fair				Fair			52.94
Study 5	Fair				Fair			47.06
Study 6	Fair				Fair			47.06
Siegling and Petrides (2016)	Fair			Good	Poor			55.56
DERS								
Gratz and Roemer (2004)								
Study 1	Fair		Poor	Fair	Poor			46.43
Study 2		Poor						45.45
S-DERS								
Lavender et al. (2017)	Fair		Poor	Fair	Fair			64.29

Measure	Internal consistency	Reliability	Content validity	Construct validity		Criterion validity	Responsiveness	Excellent Ratings (%)
				Structural validity	Hypothesis testing			
ASQ								
Hofmann and Kashdan (2010)								
Study 1	Poor		Poor	Poor	Poor			39.29
Study 2	Fair			Fair	Poor			54.55

Note. NR= not reported; AAQ = Acceptance and Action Questionnaire; AAQ-II = Acceptance and Action Questionnaire – Revised; FFMQ = Five Facet Mindfulness Questionnaire; FFMQ-15 = Five Facet Mindfulness Questionnaire-Short Form; KIMS= Kentucky Inventory of Mindfulness Skills; PHLMS= Philadelphia Mindfulness Scale; DERS= Difficulties in Emotion Regulation Scale; S-DERS= State Difficulties in Emotion Regulation Scale; ASQ= Affective Style Questionnaire,

Table 1.5

Measurement Property Quality per Instrument

Study	Content validity	Internal consistency	Criterion validity	Construct validity	Reproducibility: Reliability	Responsiveness	Interpretability	Total Positive Ratings
AAQ								
Hayes et al (2004)	-	+	0	?	?	0	?	1
Kortte et al. (2009)	0	?	0	+	0	0	0	1
AAQ-II								
Bond et al. (2011)								
Study 1	+	+	0	0	0	0	0	1
Study 2	0	+	0	0	?	0	?	1
Study 3	0	0	0	+	0	0	0	1
Palladino et al. (2013)	0	+	0	?	0	0	0	1
Scott et al. (2016)	0	?	0	0	0	0	0	1
FFMQ								
Andrei et al. (2016)	0	?	0	?	0	0	0	0
Baer et al. (2008)	0	?	0	+	0	0	?	1
Christopher et al. (2012)	0	+	0	?	0	0	0	2

Psychometric Properties of Acceptance Measures: A Systematic Review

Study	Content validity	Internal consistency	Criterion validity	Construct validity	Reproducibility: Reliability	Responsiveness	Interpretability	Total Positive Ratings
FFMQ								
Curtiss and Klemanski (2014)	0	?	0	+	0	0	0	1
Goldberg et al. (2016)	0	?	0	?	0	?	0	0
Gu et al. (2016)	0	?	0	+	0	?	0	1
Hawley et al. (2017)	0	+	0	?	0	0	?	1
Siegling and Petrides (2016)	0	?	0	?	0	0	0	0
Watson-Singleton et al. (2018)	0	-	0	?	+	0	0	0
Williams et al. (2014)	0	?	0	0	0	0	0	0
FFMQ-15								
Gu et al. (2016)	0	+	0	+	0	?	0	2
KIMS								
Baer et al. (2004)								
Study 1	-	0	0	0	0	0	0	0
Study 2	0	?	0	0	0	0	0	0
Study 3	0	0	0	0	?	0	0	0
Study 4	0	0	0	+	0	0	0	1
Study 5	0	0	0	+	0	0	0	1

Study	Content validity	Internal consistency	Criterion validity	Construct validity	Reproducibility: Reliability	Responsiveness	Interpretability	Total Positive Ratings
KIMS								
Baer et al. (2004)								
Study 6	0	0	0	?	0	0	?	1
PHLMS								
Andrei et al. (2016)								
	0	?	0	?	0	0	0	0
Cardaciotto et al. (2008)								
Study 1	-	0	0	0	0	0	0	0
Study 2	0	?	0	0	0	0	0	0
Study 3	0	+	0	+	0	0	0	2
Study 4	0	?	0	+	0	0	?	1
Study 5	0	?	0	?	0	0	?	1
Study 6	0	?	0	+	0	0	?	1
Siegling and Petrides (2016)								
	0	?	0	?	0	0	0	0
DERS								
Gratz and Roemer (2004)								
Study 1	?	+	0	?	0	0	?	1
Study 2	0	0	0	0	?	0	0	0

Psychometric Properties of Acceptance Measures: A Systematic Review

Study	Content validity	Internal consistency	Criterion validity	Construct validity	Reproducibility: Reliability	Responsiveness	Interpretability	Total Positive Ratings
S-DERS								
Lavender et al. (2017)	?	+	0	+	0	0	0	2
ASQ								
Hofmann and Kashdan (2010)								
Study 1	-	?	0	?	0	0	0	0
Study 2	0	-	0	?	0	0	0	0

Note. Rating: + ‘positive’, ? ‘indeterminate’, - ‘negative’, 0 ‘not reported/no information available’; AAQ = Acceptance and Action Questionnaire; AAQ-II = Acceptance and Action Questionnaire – Revised; FFMQ = Five Facet Mindfulness Questionnaire; FFMQ-15 = Five Facet Mindfulness Questionnaire-Short Form; KIMS= Kentucky Inventory of Mindfulness Skills; PHLMS= Philadelphia Mindfulness Scale; DERS= Difficulties in Emotion Regulation Scale; S-DERS= State Difficulties in Emotion Regulation Scale; ASQ= Affective Style Questionnaire,

1.4 Discussion

1.4.1 Summary of findings

The aim of the current review was to systematically evaluate scales which sought to measure acceptance. A comprehensive database search identified 3097 potential articles, from which 20 articles, comprising 32 sub-studies, met inclusion criteria. Nine measurement tools were identified, and these were either unidimensional measures of acceptance/psychological flexibility ($n=2$) or multidimensional tools of wider constructs which included an acceptance subscale ($n=7$). The COSMIN criteria were used to evaluate methodological quality, and we planned to evaluate all aspects of the taxonomy, with the exception of cross-cultural validity as no translated measures were included in this review. However, no studies assessed measurement error and therefore this has not been reported. Most studies assessed internal consistency and construct validity. The FFMQ was the most widely evaluated measure, followed by the AAQ-II and the PHLMS.

1.4.2 Theoretical definitions of acceptance

Throughout this review, many of the tools being evaluated were used as evidence of convergent validity for one another, despite no established ‘gold standard’ of acceptance and a lack of evidence for the optimal measure. While some studies showed that (sub)scales of acceptance loaded onto the same factor or shared variance, the amount of shared variance and relationship strength varied across studies, calling into question what is being measured by each tool. For example, the AAQ was developed as a measure of acceptance/experiential avoidance which were originally conceptualised as the main focus of ACT. However, as ACT broadened its definitions and understanding, these terms were seen as examples of psychological flexibility/inflexibility rather than the representing the

construct as a whole. This shift in language and understanding has not necessarily percolated through the literature, and despite the fact that the AAQ-II also measures other aspects of the hexaflex, such as committed action, it has continued to be associated with acceptance. This may partially account for less shared variance and weaker relationships with other scales of acceptance. However, there are also significant implications for development of other instruments, as the AAQ-II is often treated as the de facto 'gold standard' acceptance measure.

Another consideration is the underlying theoretical approaches from which the measures were developed. In this review, the majority of scales formed part of mindfulness tools, two others were developed within ACT, and three more were developed from an emotion regulation perspective. Thus, how acceptance is conceptualised varies across approaches, and this in turn influences what is measured. However, the underlying factor structure of items, and whether these map onto shared latent variables is yet to be explored.

As identified throughout this review, defining and measuring acceptance is difficult. Thus, language and meaning have significant implications, especially for those not familiar with the concept. For example, the authors of the ASQ describe an intention to measure an accepting attitude of emotional experience, but they label this 'tolerating' (a word that appears in several items of the scale). The word tolerating may convey a more passive construct than the conceptualisation of acceptance in ACT, which is both active and dynamic. Thus, while the scale claims to measure a construct synonymous with acceptance, it is not clear whether this semantic interpretation is shared by participants due to the language used in the scale, and lack of target population involvement at the item development stage.

To a great extent, all of the above difficulties described above result directly from shortcomings in content validity. Despite acceptance being considered a nuanced concept,

as noted in this review, there was often no involvement with target populations for item generation and evaluation. Thus, how non-experts understand and interpret items across scales is not fully understood, nor is the influence of patient demographics upon the semantic meaning of items. Qualitative approaches to understanding the relevance, comprehensiveness and understanding amongst target populations are needed to explore this further.

1.4.3 Acceptance and Non-acceptance

Another important consideration throughout this review was understanding the dimensionality of acceptance. Two of the scales measured non-acceptance using negatively worded items (DERS and S-DERS), while all of the items on the FFMQ, KIMS and PHLMS are negatively worded and then reversed scored to measure acceptance. The AAQ and AAQ-II also measure experiential avoidance/ psychological inflexibility and suggest that lower scores on these constructs indicate greater acceptance/ flexibility. In doing so, such scales assume that acceptance and non-acceptance are polarities of the same construct. However, this might not be the case. For example, when developing the Multidimensional Experiential Avoidance Questionnaire (MEAQ), Gámez, Chmielewski, Kotov, Ruggero, and Watson (2011) found that items pertaining to willingness (an important part of acceptance), did not negatively load onto a non-acceptance factor, but instead loaded onto a distinct factor. Thus, much like positive and negative affect are no longer considered as polarities of an underlying construct, but instead seen as distinct (albeit related) dimensions, this too might apply when considering acceptance and non-acceptance. If this were the case, there would be significant implications for the validity of existing measures which use inverse scoring. Further research is needed to fully understand the dimensionality of acceptance to inform suitable measures of this important construct.

1.4.4 Utility of Acceptance Scales

This review suggests that current evidence is not sufficiently comprehensive or robust to guide the choice of suitable acceptance measures. Many measurement properties were not adequately assessed or reported, or were absent completely (such as floor and ceiling effects, and measurement error). Few measures were assessed with regards to test-retest reliability and responsiveness, and those that did were of poor quality, often with small sample sizes, poorly detailed rationales for follow up time frames or suboptimal statistical analyses. Furthermore, in theoretical domains such as ACT and third-wave therapies, it remains unclear which is the best measure. Few scales have been developed within this framework, and studies have relied on the AAQ-II. However, the AAQ-II has been criticised for conflating process with outcome, as many items in the scale also refer to affect experienced (Wolgast, 2014).

1.4.5 Use of COSMIN and Terwee Quality Criteria

Almost all studies which evaluated content validity were given a poor rating, with the exception of Bond et al. (2011). Both quality criteria used in the current review require the involvement of target populations in assessment of item relevance in order to receive a positive rating. However, unlike more traditional health related outcome measures, many of the mindfulness, emotion regulation and acceptance tools are intended to be used across a range of clinical and non-clinical populations. Consequently, authors rarely identify a target population or only make indirect reference to the purpose of the scale. As a result, item generation and selection was often undertaken solely by experts, which resulted in lower ratings. As already mentioned in this review, the paucity of target population or non-expert involvement could have significant implications for the accuracy of acceptance measures, and limit any conclusions that can be drawn about their utility in specific populations.

Throughout this review, many studies conducted multiple correlations in assessments to demonstrate convergent, discriminant and concurrent validity. However, few reported systematic attempts to compensate for type I errors. Only two articles, Gu et al. (2016) and Kortte et al. (2009), used Bonferroni corrections, while Curtiss and Klemanski (2014) was the only study to use FDR procedures to adjust p values. Two studies arbitrarily changed the significance level to $p < .01$ (Baer et al., 2004; Cardaciotto et al., 2008). It has been suggested that Bonferroni corrections with large number of comparisons are overly conservative, and as such attempts to avoid type I errors are at the cost of increased likelihood of type II errors (Perneger, 1998). Methods such as FDR have therefore been proposed as a viable and robust alternative (Benjamini & Yekutieli, 2001). To reduce overdependence on null-hypothesis testing and consider results in a wider context, it is now recommended that confidence intervals and effect sizes are reported in addition to significance values, so as to determine whether the identified values are likely to fall within the true range (Greenland et al., 2016). While this relatively new reporting convention has come about since the publication of many of the articles reviewed here, it is an important consideration for future research. In the present review, studies which demonstrated attempts to reduce type I errors, but which had not done so in a systematic way, were rated as having minor methodological problems, whereas those who made no attempts to adjust for multiple comparisons were rated as having important methodological flaws.

The quality criteria in this review state explicit requirements for factor analyses to ensure robustness, such as sample size to item ratio. The criteria also require suitability of analysis methods, EFA or CFA. However, no criteria with regards to the model fit estimates are proposed, and as such, the quality of accepted models across studies might vary considerably. Furthermore, a number of studies employed parcelling methods as a way to circumvent samples that were smaller than desired without losing power in the

model. Parcelling, an approach whereby items are combined together into parcels, is the source of much debate, especially within scale development. Different methods of parcelling exist each of which has relative benefits and drawbacks and therefore parcelling should be used with caution (see Little, Rhemtulla, Gibson, & Schoemann, 2013 for a review). This is especially true in scale development where true relationships amongst individual scale items might be hidden by combining them into parcels and so the use of parcels is rarely warranted (Little et al., 2013; Worthington & Whittaker, 2006). Given the controversy over parcelling, and the lack of guidance for rating this in the quality criteria, studies were not rated down for using this method and the stipulated sample-to-items ratio was used across all studies conducting factor analyses, irrespective of parcelling. However, future reviews may wish to consider rating criteria for item parcelling.

The COSMIN criteria place great emphasis on how missing data is handled, and this is rated in almost every element of the taxonomy. Therefore, studies which do not report this process are unable to obtain more than a ‘fair’ rating. While missing data is an important consideration which can lead to biases, discrepancies in reporting and practice might result in a negative skew in scoring (i.e. some authors may have managed missing data well, but not reported this in their publication). Some reviews have modified the use of COSMIN to allow studies to score higher if missing data was the only flaw in a particular section to compensate for potential bias (Park, Reilly-Spong & Gross, 2013). This adjustment was not made in the present review. Instead, the authors evaluated the percentage of excellent ratings studies achieved across measurement properties, which allowed comparison of studies with generally lower quality ratings to those which might have been negatively biased by the “worst score counts” rating system. This process permitted identification of when overall scores for measurement properties were significantly affected by a single score, or repeatedly affected by the same issue.

1.4.6 Strengths

A strength of the current review was that it employed two well established assessment criteria to assess both methodological and measurement property quality. The use of established assessment tools ensures robust assessment of instruments, and enables comparisons of findings across reviews. The excellent inter-rater reliability in the present study also demonstrates the benefits of using such established tools.

1.4.7 Limitations

There are a number of limitations to the present review which much be considered. Firstly, only one reviewer conducted the first stage review of over 3000 article titles and abstracts which might have resulted in bias. However, a second rater was involved in full paper selection and quality ratings, which demonstrated excellent agreement. Secondly, there is also a risk of selection bias by only including studies and measures in English. This decision also precluded any analysis of cross-cultural validity. However, given the significant number of translated measures available, a separate review is warranted to address their validity. Thirdly, it is possible the search criteria employed were overly stringent, meaning potential studies and measures were not identified. Some work has been done on developing a standardised search, but as yet this is only available on certain platforms and so it was not possible to use in this study (Terwee, Jansma, Riphagen, & de Vet, 2009).

Measures which were designed to measure acceptance of a particular concept or in a given context, were also not included in the current study. Given that these scales aim to measure a very particular type of acceptance, further research is needed to explore how these measures relate to the general process of acceptance.

COSMIN and Terwee et al.'s (2007) quality criteria are relatively new and could be considered quite restrictive. As such, advances in reporting will not be matched in older studies which also will not have used such criteria to guide their work. Thus, there is a risk of negative bias for studies which do not report heavily weighted criteria such as missing data handling. However, future studies would benefit from using these criteria to guide reporting of key measurement and methodological properties.

1.4.8 Conclusion

To conclude, self-report measures of acceptance have the potential to be highly valuable in assessing mechanisms and outcome of therapies which emphasise the importance of this process. Further research is needed to establish whether existing measures developed from different therapeutic models are assessing the same latent construct. Additionally, research is required to understand the dimensionality of acceptance and non-acceptance, the results of which will have implications for how each construct is assessed. Content validity was considered to be lacking across measures. Further research, including diverse target populations, is required to ensure comprehensiveness, relevance and clarity of items across scales. Given the implications for other forms of validity and reliability across populations, this research should take precedence. Whilst the stringency of the quality criteria might have resulted in a negative bias, as evidenced by the percentage of excellent ratings, it nevertheless provides robust and systematic evaluation of measurement properties which can inform future research in this important area. As more therapeutic approaches begin to emphasise the role of acceptance in coping with difficulties, the need for reliable and valid measures increases in both clinical and research settings. With this in mind, it is hoped that this review can provide some guidance as to the quality of existing tools, and inform future research into the psychometric properties of acceptance measures.

Chapter 2 Empirical Paper: Development of the Southampton Acceptance Scale (SAS)

2.1 Introduction

In recent years, a growing number of psychological therapies have emphasised acceptance as a crucial process for psychological wellbeing. However, the definition of acceptance varies across different branches of psychological literature and measures of acceptance have traditionally been developed within a particular theoretical school. There is a paucity of reliable and valid measures that are theoretically coherent with acceptance-based psychological therapies, and this has implications for the evaluation of such treatments in both clinical and research settings. The aim of the current paper is to develop a new measure of acceptance which is theoretically consistent with acceptance-based psychological approaches.

2.1.1 Definitions of Acceptance

Acceptance is a term used widely throughout various domains of psychological literature, including coping, emotion (dys)regulation and psychological therapies. While the same term is employed, the constructs under consideration vary in important ways. Each of these definitions will be discussed, as will the relative similarities and differences between them.

Coping is the process of changing one's cognitive and behavioural responses (also known as stress reactions) to meet or manage demands following threat appraisal (Lazarus & Folkman, 1984). The coping literature defines two general types of coping, problem-focused and emotion-focused. Problem-focused coping seeks to alter the environment or the source of the stress, whereas emotion-focused coping is concerned with managing the

Development and initial validation of the Southampton Acceptance Scale (SAS)

emotional distress experienced in response to the situation. In the coping literature acceptance is considered as an emotion-focused coping strategy, and describes the process of coming to terms with, and moving on from, adverse life events and experiences. It is conceptualised as the opposite of denial which is considered as a maladaptive coping strategy (Carver, Scheier, & Weintraub, 1989).

Emotion regulation is the process of modulating emotional experiences. Emotion regulation describes the enduring process present throughout everyday life, as individuals up- and down-regulate their responses to emotionally arousing stimuli (Kooze, 2009). Within the literature, significant attention has been paid to emotion regulation difficulties (dysregulation) and their role in psychopathology. Emotion dysregulation is defined as difficulty coping with emotions, and can pertain to how emotions are experienced or processed (Leahy et al., 2011). Dysregulation can result in intensification of emotions, so that the emotion is perceived to be overwhelming and intolerable. Conversely, dysregulation can also result in deactivation of emotions, including emotional numbing, depersonalisation, and dissociation. In some contexts these emotional responses can be functional or even adaptive in the short term, such as during a traumatic event to enable coping (Leahy et al., 2011). However, when intensification or deactivation occurs over longer periods of time, or in contexts where there is no functional benefit, then these strategies are considered to be maladaptive, and thus indicative of dysregulation. Emotion dysregulation has been identified as a core component in complex mental health difficulties, such as Borderline Personality Disorder (BPD; Linehan, 1993) and complex Post-traumatic Stress Disorder (PTSD; Cloitre et al., 2009). Given that, in these conditions, emotions are often experienced as overwhelming, or excessively suppressed, they are often considered to be intolerable and therefore negative. This non-acceptance of emotions is considered to be maladaptive and a key feature of difficulties such as BPD (Linehan, 1993). Therefore, in the emotion regulation literature acceptance is often conceptualised in

terms of emotional acceptance; accepting and valuing one's emotional responses (Gratz & Roemer, 2004).

Within "third-wave" cognitive-behavioural therapies, acceptance describes the process of experiencing private events, such as thoughts and emotions, for what they are (Hayes, 2004). In this context, acceptance is an active process, that involves being willing and open to experience all private events, non-judgementally, without engaging in attempts to change them.

All three areas of literature recognise the significant role of acceptance for psychological wellbeing and affect. However, there are also important distinctions between these conceptualisations of acceptance and their function. Firstly, acceptance as defined within third-wave therapies is an active process which applies to all internal private experiences, whereas the coping literature is predominantly concerned with coming to terms with (often external) stressors. Secondly, acceptance-based coping seeks to reduce the psychological distress associated with a stressor, and acceptance in the context of emotion regulation aims to modulate the emotional response. Conversely, acceptance in third-wave therapies does not seek to reduce emotional experience or intensity, but rather to enable valued living (Blackledge & Hayes, 2001). Acceptance within the emotion regulation literature is a secondary process which shares some overlap with acceptance as defined in third-wave approaches, but is a much narrower definition; it does not consider acceptance of other private events, such as thoughts and physiological sensations.

Given these important distinctions between conceptualisations of acceptance, it is necessary for researchers to be clear which construct is under consideration. As such, this report will focus on the construct of acceptance as defined in third-wave psychological therapies, and particularly within Acceptance and Commitment Therapy (ACT; Hayes,

Development and initial validation of the Southampton Acceptance Scale (SAS)

Strosahl, & Wilson, 1999). An overview of the ACT model follows, together with a summary of the importance of acceptance within this framework.

2.1.2 Acceptance and Commitment Therapy

Second generation cognitive behavioural therapies propose that psychological distress occurs as a result of maladaptive interactions between cognitions, behaviours, emotions, and physiology. Direct cognitive change is a central pillar of the therapy and is considered a requisite for clinical improvement and reduced distress (Westbrook, 2014). Third-wave psychological therapies on the other hand advocate changing one's relationship to psychological experiences and to their function (Hayes, Levin, Plumb-Villardaga, Villatte, & Pistorello, 2013). ACT (Hayes et al., 1999) is a transdiagnostic "third-wave" cognitive-behavioural therapy grounded in relational frame theory (RFT; Hayes, 2004), which emphasises the importance of language and cognition in psychopathology. RFT suggests that through language humans derive and infer meaning and relationships between different pieces of information. Through this process, emotional content becomes associated with certain objects, words, memories etc., which can lead to such stimuli being experienced as aversive. As such, both the stimulus itself and thoughts about this without the actual presence of the stimuli can lead to distress. This often results in attempts to avoid experiencing the stimuli, either physically or cognitively so as to reduce distress. As more associations are then made with the distressing stimuli, and the relational frame expands, more stimuli are avoided due to perceived threat. The role of language and inferred relationships between information and how these relational frames grow to include more information can be conceptualised as the basis for many mental health difficulties.

Based on RFT, the ACT model of psychological distress posits that over-dominance of cognitive processes result in attempts to avoid, escape or control unwanted and aversive

private experiences. Known as experiential avoidance, these avoidance attempts often come with behavioural and emotional costs; individuals are more likely to prioritise the avoidance of unpleasant experiences over the pursuit of value-driven goals and behaviours (Hayes et al., 1999). Consequently, in the long term, experiential avoidance and the absence of value-driven action results in greater psychological suffering. Psychological inflexibility is the term used in ACT to describe this interaction between cognitive and behavioural processes, and is considered to be the core of most human suffering and psychopathology (Hayes et al., 2013).

ACT therefore seeks to reduce experiential avoidance and promote value-driven behaviours through cultivating greater psychological flexibility, defined as “*the ability to contact the present moment more fully as a conscious human being, and to change or persist in behavior when doing so serves valued ends*” (Hayes, Luoma, Bond, Masuda, & Lillis, 2006, p.7). The ACT model specifies six key processes, considered as psychological skills, which foster psychological flexibility. These processes are often represented using the ‘Hexaflex’ model (Harris, 2009), which visually represents each of these six processes and how they are interconnected. Acceptance, unsurprisingly given the name of the therapy, is an integral process in ACT.

2.1.3 Acceptance and ACT

The ACT model proposed that each psychological process has a corresponding pathological process, and these are polarities of the same underlying construct. Within ACT, acceptance is considered as the antithesis of experiential avoidance; it is the process of openness and willingness to embrace private experiences for what they are, without engaging in attempts to alter the form, frequency or intensity of them (Hayes et al., 2006). Acceptance is an active, intentional behaviour that seeks to alter the function of private experiences from a threat to be avoided, to a part of valued living to be observed and

Development and initial validation of the Southampton Acceptance Scale (SAS)

approached with curiosity and openness (Hayes et al., 2013). Acceptance in this context is distinguished from the passive processes of resignation and tolerance. It is also not tantamount to liking or condoning the experience; it is possible to dislike the experience and still accept that this is the present reality. Furthermore, ACT does not advocate acceptance of all experiences, only those that are not readily changeable or where non-acceptance would incur behavioural costs. As such, acceptance is not considered an end goal, but rather something to be fostered in the service of living a valued and meaningful life (Blackledge & Hayes, 2001).

The definition of acceptance describes how to be in relation to experiential events, providing a general indication of how to approach these events. However, the specific processes of how this is achieved is less clear. For example, adopting a position of openness and willingness to experiences is a central feature of acceptance. However, this is a rather abstract concept which can be difficult to operationalise; the specific cognitive and emotional processes that facilitate this position are less well understood. These difficulties in operationalising processes associated with acceptance have made it a difficult concept to measure.

2.1.4 Measuring acceptance

Given that acceptance is a term used to describe several different constructs, the first challenge for researchers is to identify a tool which measures the type of acceptance that is of interest. Within the ACT literature there are a number of tools that aim to measure the key ACT processes, including acceptance. However, each of these has limitations when seeking to measure acceptance specifically.

2.1.4.1 General versus specific measures of acceptance

Several measures have been developed to assess the six key processes of the ACT model. Perhaps the most well-known and widely used ACT measure is the Acceptance and Action Questionnaire-Revised (AAQ-II; Bond et al., 2011), which assesses psychological inflexibility. This seven-item scale seeks to measure various aspects of the ACT hexaflex, with a particular emphasis on experiential avoidance and fusion. However, many items conflate several processes within a single question, and it is difficult to separate out key processes. A more recent measure is the Comprehensive Assessment of Acceptance and Commitment Therapy processes (CompACT; Francis, Dawson, & Golijani-Moghaddam, 2016). This scale conceptualises the ACT model in terms of three dyadic processes, and so measures both acceptance and defusion on a single subscale. Consequently, while useful for reducing respondent burden and assessing the ACT model as a whole, these measures lack specificity. Furthermore, it is not possible to calculate scores for independent processes, meaning they are ill equipped to measure acceptance as a standalone process.

Regarding specific measures, there has been a proliferation in tools seeking to measure acceptance of specific mental and physical health difficulties, including voices (Shawyer et al., 2007), body image (Sandoz, Wilson, Merwin, & Kate Kellum, 2013), chronic pain (McCracken, Vowles, & Eccleston, 2004) and tinnitus (Westin, Hayes, & Andersson, 2008). However, there are a number of limitations to these measures when seeking to measure acceptance more generally. Firstly, rather than looking at experiential acceptance as a general process, these tools focus on acceptance of a particular construct. For example, items on the chronic pain acceptance questionnaire (CPAQ; McCracken et al., 2004) ask specifically about the impact of chronic pain and how the individual manages thoughts and feelings related to pain. As such, these measures are not suitable for use with populations beyond those for which they are designed. Secondly, many of these scales were developed by adapting items from the AAQ (Hayes et al., 2004) and as such

Development and initial validation of the Southampton Acceptance Scale (SAS)

are likely to measure multiple processes of the ACT model, so are not solely measuring acceptance. Furthermore, despite using the AAQ as a starting point to generate items, the scales contain different items, which means that they cannot directly be compared.

Therefore, it is not possible to consider whether acceptance of a particular difficulty experienced by a specific population relies on the same processes as more general experiential acceptance. Consequently, these measures lack generalisability, and as they do not measure the process of acceptance per se, their application is limited.

Presently, there are a number of ACT process-specific measures which pertain to particular aspects of the Hexaflex, such as the Cognitive Fusion Questionnaire (CFQ; Gillanders et al., 2014), and the Committed Action Questionnaire (CAQ; McCracken, 2013). However, there is no identified measure of acceptance as conceptualised within the ACT model. While the AAQ-II is a measure of psychological inflexibility, its predecessor, the Acceptance and Action Questionnaire (AAQ; Hayes et al., 2004) was defined as a measure of acceptance and experiential avoidance before the concepts were redefined as psychological flexibility and inflexibility respectively. Consequently, although a general measure of the ACT model, the AAQ-II is often used as a measure of acceptance in the absence of other available tools.

2.1.4.2 Other acceptance measures

Beyond ACT-specific measures, a number of tools contain acceptance and non-acceptance scales, including measures of mindfulness (e.g., Five Facet Mindfulness Questionnaire; Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006), and emotion regulation (e.g., Difficulties in Emotion Regulation Scale; Gratz & Roemer, 2004). While there appears to be some overlap in the constructs measured by these scales, there is as yet insufficient evidence as to whether they target the same latent variables to warrant their use

as ACT process measures. Furthermore, many of these scales rely on reverse scored items and as such may share the dimensionality concerns identified above.

2.1.4.3 Process versus outcomes

Given that acceptance and experiential avoidance predict emotional experience, the two constructs are strongly related. However, it is important that measurement tools can distinguish between the process of responding to private events and the subsequent emotional experience. The AAQ-II has been criticised for its failure to distinguish outcome from process. During development, the AAQ-II was shown to be highly correlated with a depression measure, so the authors performed a confirmatory factor analysis (CFA) to determine whether the items of the two scales were better explained by a single factor, or two distinct factors. The analysis showed a two-factor solution was a better fit to the data, suggesting the two scales were not measuring the same latent variable (Bond et al., 2011). While the initial factor analyses demonstrated sufficient concurrent validity, Wolgast (2014) suggests that the focus on distress within items may result in multiple interpretations, thus confounding process with outcome. Francis et al. (2016) also found strong cross-loadings of AAQ-II items across factors measuring openness to experience (acceptance and defusion) and distress. Similarly, Gámez, Chmielewski, Kotov, Ruggero, and Watson (2011) have argued that AAQ-II's emphasis on distress leading to dysfunction may result in item contamination, making it unclear whether responses reflect levels of experiential avoidance, levels of distress, or a combination of the two.

As previously stated, when developing the AAQ-II, Bond et al. (2011) conducted a CFA with the AAQ-II and a measure of depression, comparing a one- and two-factor solution. In the one factor model, items from both scales were loaded onto a single latent variable, and the fit of this model was compared to the fit of a two-factor solution, where the AAQ-II and depression were identified as distinct, but related, latent variables. The two

Development and initial validation of the Southampton Acceptance Scale (SAS)

factor model was demonstrated to be a better fit overall. The authors claim that this demonstrated concurrent validity of the two scales. However, this approach did not allow for items to cross-load because CFA relies on a priori hypothesis of relations between items and latent variables. Thus, individual items are not free to load onto factors in the same way as exploratory factor analyses (EFA). Furthermore, while the overall model was a better fit to the data, neither model was a good fit, and the authors do not report whether modification indices recommended any cross loadings be added to the model to improve fit. As such, it is unclear whether particular items on the AAQ-II demonstrate greater shared variance and strong loading on the depression factor. Furthermore, depression is only one possible form of psychological distress, as such the overlap of AAQ-II items with general psychological distress remains unclear.

2.1.4.4 Dimensionality of acceptance and experiential avoidance

As previously mentioned acceptance is the counter process to experiential avoidance and within the ACT model these two processes are considered to be opposing poles of the same underlying construct. However, during the development of the Multidimensional Experiential Avoidance Questionnaire (MEAQ) Gámez and colleagues (2011) identified that reverse-scored acceptance items did not directly load onto non-acceptance (experiential avoidance) scales, but instead loaded onto a distinct factor. This would suggest that acceptance and experiential avoidance are potentially distinct, albeit related, constructs. As there is not presently an independent measure of acceptance, this hypothesis has not yet been tested. However, if this were to be the case, it would have significant implications for existing scales, whose items are almost exclusively focused on non-acceptance and then reverse scored.

2.1.4.5 Summary of limitations of current measures

While a number of existing tools seek to measure acceptance, there are a number of limitations with these measures. Firstly, many tools were developed outside an ACT framework, and so it is unclear whether these measure the same construct. Of those tools developed within ACT, several measure multiple components of the ACT model and do not provide acceptance scores in isolation. Other instruments measure acceptance of a given construct rather than the process itself. Finally, many tools measure experiential avoidance using reverse score items, which raises concerns over their validity given uncertainty surrounding the dimensionality of these constructs.

2.1.5 Research Objectives

The lack of a valid and reliable measure of acceptance prevents researchers from drawing conclusions about the efficacy of acceptance-based interventions and identifying the role of acceptance processes in clinical change. Given this significant limitation, the aim of this paper was to develop and evaluate a new measure of acceptance which accurately reflects the construct as conceptualised within ACT. Furthermore, we also sought to investigate the dimensionality of acceptance and experiential avoidance.

We conducted three separate studies to develop the scale and explore specific psychometric properties. Specifically, we sought to evaluate the factor structure, reliability, and validity of this new measure. An initial exploratory approach was employed as a unique pool of items with many potential relationships was under investigation. These analyses were subsequently followed up with confirmatory approaches.

For the purposes of the new measure, acceptance was defined as the process of “*actively contacting psychological experiences – directly, fully, and without needless defense*” (Hayes, Wilson, Gifford, Follette, & Strosahl, 1996, p. 1163). In study one, an

Development and initial validation of the Southampton Acceptance Scale (SAS)

initial item pool was generated and assessed by a panel of experts, and reviewed by non-experts. Study two sought to refine the item pool into a useable scale and identify the factor structure, using exploratory factor analyses. In study three the proposed factor structure was tested using a confirmatory approach. Once this was established, the construct, convergent, and divergent validity of the measure was assessed. Internal consistency was assessed across all three studies. This new measure was named the Southampton Acceptance Scale (SAS).

2.2 Phase One

2.2.1 Method

The University of Southampton Research Ethics Committee granted ethical approval for all phases of the studies, which were undertaken as part of a wider project on emotion regulation (Appendix D).

2.2.1.1 Preliminary item pool.

2.2.1.1.1 Operational definition of acceptance

As stated above, in the present study acceptance was defined as the active process of fully and directly contacting psychological experiences, without needless defence (Hayes et al., 1996). Acceptance was distinguished from passive tolerance, and defined as an active process of embracing experiences as they are now (Hayes, 2004). In other words, the process by which individuals actively make room for thoughts, feelings, and memories to come and go naturally, without trying to control or change them, to take them as they are, even if they are unpleasant. This definition was used to guide the development and selection of items.

2.2.1.1.2 Item pool development.

An initial pool of 120 items was generated by the main author based on a combination of theoretical understanding, literature pertaining to acceptance, and reviewing existing measures of acceptance and psychological flexibility (Bond et al., 2011; Cardaciotto, Herbert, Forman, Moitra, & Farrow, 2008; Francis et al., 2016). To ensure broad applicability, items were generated that focused on a range of internal and private events, including thoughts, memories, feelings, and emotions. Furthermore, given that acceptance of negative private events is considered to be a more difficult process, items were developed to expressly ask about unpleasant thoughts and emotions in addition to thoughts and emotions more generally. Positive and negatively worded items were included to reduce acquiescent responding (DeVellis, 2003) and to explore dimensionality, as it was unclear how these items would load onto factors. Negatively worded items were added in two ways, by changing the orientation of the question, and by reversing the meaning of items. Regarding reverse orientation, items included negatively cued words, such as ‘not’ or ‘unpleasant’, for example “*I allow myself to experience unpleasant thoughts even if they get worse*” and “*I do not try to control my thoughts, even if they are unpleasant*”. Reverse meaning items were used to ask the opposite of one another, such as “*It is important to control my emotions*” and “*I am willing to fully experience all emotions that arise*”.

In line with recommendations for scale development (DeVellis, 2003), the initial item pool was distributed amongst ACT experts for verification and feedback. Experts were defined as clinicians or researchers who met one or more of the following criteria: i) five or more years using ACT as a primary model in clinical practice, ii) provided teaching or training in ACT, or iii) had two or more ACT-related publications. Experts were recruited through local NHS Trusts, and through social media special interest sites for ACT practitioners and researchers. Experts were provided with the definition of acceptance that

Development and initial validation of the Southampton Acceptance Scale (SAS)

had been used to develop items, and were then invited to rate each item of how closely the item related to this definition. Items were rated on a 10-point Likert scale, from 1 (*'not at all'*) to 10 (*'completely describes this'*). Experts were also invited to give qualitative feedback regarding clarity, conciseness, acceptability, or any other aspects of the item they felt were important or relevant. In line with best practice, consensus criteria were defined a priori (Diamond et al., 2014). As such, ratings of seven or above were to be considered as highly congruent with the definition. Consensus was predefined as ratings of high congruence by 75% or more of experts.

Furthermore, as the scale aims to measure acceptance in a wide range of individuals in clinical and non-clinical populations who might not be familiar with ACT, an opportunity sample of personal contacts who were not psychologists or familiar with ACT were invited to give qualitative feedback about the clarity, accessibility, and understanding of items.

2.2.1.1.3 Participants

Five ACT experts were recruited to review the item pool in line with recommendations for establishing content and face validity (Netemeyer, Bearden, & Sharma, 2003). Participants were asked to read the participant information sheet before agreeing to take part (Appendix C). All experts were Clinical Psychologists, and provided teaching or training in ACT, with at least one year's clinical experience of using ACT as a primary model. Two participants also had ACT-related publications. The mean age of the sample was 37 years ($SD = 9.03$, range 29-52), and 60% of the sample were female.

Non-experts comprised of five adults who were unfamiliar with ACT. The mean age of the sample was 37 ($SD = 13.76$, range 23-54), and 40% of the sample were female.

2.2.2 Results and discussion

Based on the predefined consensus criteria, items with strong inter-rater agreement of high congruence (75% or greater) amongst experts were retained, which resulted in a fifty-two item scale.

Qualitative feedback by experts and non-experts was reviewed, and items were amended or removed as recommended and in line with the definition. Specifically, feedback highlighted items which discussed “feelings” were ambiguous as to whether this referred to emotions or physiological sensations. As such, several items were reworded to describe emotions specifically. The retained items are summarised in Table 2.1.

2.3 Phase Two

Study two sought to examine the factor structure of the retained 52-items and reduce these down into a theoretically consistent and practical scale which could be used in clinical and research settings. In line with ethical considerations, all participants were provided with an information sheet which outlined the purpose of the study, data handling and the right to withdraw at any time (Appendix E) prior to agreeing to take part. After completing the study, participants were provided with a debrief sheet which also contained information about further support available should participants have experienced any negative emotions as a result of taking part (Appendix F).

Development and initial validation of the Southampton Acceptance Scale (SAS)

Table 2.1

Initial Exploratory Factor Analysis Factor Loadings Across a Non-Clinical Sample (n=268)

Item	Factor one	Factor two
1. I am willing to fully experience all emotions that arise	.806	
2. I am willing to fully experience all thoughts that come to mind	.794	
3. I welcome all of my thoughts and feelings, even if they are unpleasant	.764	
4. I welcome all emotions	.750	
5. I allow my thoughts to come and go freely	.744	
6. I allow myself to experience all thoughts, even if I don't like them	.740	
7. I take thoughts and feelings as they are	.738	
8. I allow myself to experience emotions for as long as they remain	.721	
9. I allow myself to experience unpleasant thoughts even if they get worse	.713	
10. I allow myself to experience all of my emotions	.712	
11. I allow myself to experience unpleasant emotions when they arise	.706	
12. I allow myself to experience unpleasant emotions even if they get worse	.704	
13. When I feel worried or anxious, I allow myself to experience these feelings	.704	
14. I allow myself to experience unpleasant emotions, no matter how long they last	.679	
15. I allow my thoughts to be there for as long as they remain	.675	
16. I am willing to fully experience all memories that come to mind	.672	
17. I allow myself to experience unpleasant emotions	.660	
18. When I feel sad or depressed, I allow myself to experience these feelings	.635	
19. I allow myself to experience unpleasant thoughts without wishing them away	.616	
20. I do not try to control my thoughts, even if they are unpleasant	.608	
21. I welcome all of my emotions	.605	

Item	Factor one	Factor two
22. I allow myself to experience unpleasant emotions without wishing them away	.601	
23. When I feel sad or depressed, I do not try to get rid of these feelings	.536	.331
24. When I experience unpleasant emotions, I do not try to get rid of them	.506	.338
25. When I feel worried or anxious, I do not try to get rid of these feelings	.463	.386
26. It is okay to have unpleasant thoughts sometimes	.445	
27. When I feel sad or depressed, I do not try to change these feelings	.436	.315
28. I am willing to experience unpleasant feelings	.408	
29. I do not try to control my thoughts	.407	
30. It is okay to experience unpleasant emotions sometimes	.377	
31. I work hard to keep out unpleasant thoughts		.825
32. I often do things to avoid unpleasant emotions		.788
33. I work hard to avoid unpleasant feelings		.788
34. I work hard to keep out unpleasant thoughts		.727
35. When I experience unpleasant emotions, I try to change or get rid of them		.714
36. I try to distract myself when I feel unpleasant emotions		.711
37. I try to put problems out of my mind		.695
38. It is important to keep emotions under control		.695
39. I often take steps to avoid uncomfortable emotions		.672
40. It is important to control my emotions		.638
41. I tell myself I should not feel certain things		.635
42. I try to control unpleasant emotions		.625
43. I try to control unpleasant thoughts		.603
44. I try to control unpleasant thoughts and memories		.575
45. I tell myself that I should not have certain thoughts		.532
46. I try to push away unpleasant thoughts and memories		.512
47. When I feel anxious or worried, I do not try to change these feelings	.387	.397
48. I do not try to control my feelings	.331	.377

Note. Item loadings below .3 were suppressed, and so 4 items did not load satisfactorily onto either factor

2.3.1 Method

2.3.1.1 Participants.

An online independent sample of 268 participants (215 female, 2 gender fluid, 1 unidentified) was recruited through the university student research portal, social media and online research participation interest groups (<http://www.onlinepsychresearch.co.uk/>, <https://psych.hanover.edu/research/exponnet.html>, <http://www.in-mind.org/content/online-research>). The mean age of the sample was 27.81 years ($SD = 11.86$, range 18-73), and they were from United Kingdom ($n=225$, 83.7%), United States of America ($n = 24$, 8.9%), Canada ($n = 4$, 1.5%), Asia ($n=4$, 1.5%), Ireland ($n=3$, 1.1%), Australia ($n=3$, 1.1%), Europe ($n=2$, 0.7%), or not stated ($n=1$, 0.7%).

2.3.1.2 Materials and procedure.

Participants completed the 52-item pool online without compensation, with the exception of psychology students at the University who received research participation credits. After providing informed consent to take part, participants completed demographic information and then rated the items using a 7-point Likert scale from 1 ('*never true*') to 7 ('*always true*').

2.3.2 Results and discussion

2.3.2.1 Factor Structure

Exploration of data indicated some items contained missing data. Fewer than 5% of cases contained missing values, and as no pattern was evident, data was assumed to be missing at random. Cases were excluded on an analysis by analysis basis (Little & Rubin, 2014). Observation of the correlation matrix indicated that all items demonstrated at least one correlation in excess of .30 and so it was possible to proceed with factor analysis (Tabachnick & Fidell, 2014). Principal Axis Factoring with Oblique rotation was used to

explore the factor structure of the item pool. Oblique rotation was chosen as some overlap in the variance explained by identified factors was anticipated given that acceptance and experiential avoidance are not independent from one another, and this rotation permits rotated factors to correlate (Russell, 2002). Loadings below .30 were suppressed to ensure items explained sufficient variance in the latent variable (Tabachnick & Fidell, 2014). Exploration of scree plots and tables demonstrated the potential for a two or four factor structure. The two-factor structure explained 47.06% of the variance in the data, while the four factor structure explained 55.39%. However, in the four-factor solution, factors three and four explained less than 8% of the variance between them. From examination of specific items, the loading pattern and strength of these across factors, the two factor structure appeared most coherent with theory. These two factors were congruent with acceptance and experiential avoidance, and were respectively labelled “openness and willingness” (OW) which explained 36.55% of the variance, and “avoidance and control” (AC) which explained 10.51% of the variance. Internal consistency analyses were completed using Cronbach’s alpha, which was $\alpha = .95$ for OW, and $\alpha = .92$ for AC. The factor loadings of the initial EFA are summarised in Table 2.1.

Given that there was repetition and redundancy in the item pool, a repeated process by which items were removed based on theory and low factor loadings was undertaken, followed by reassessment of factor structure and internal consistency. This iterative process, in conjunction with recommendations for optimum scale length (8-10 items per subscale; Netemeyer et al., 2003) culminated in an 18-item questionnaire, comprising two subscales of 9-items each (factor loadings are summarised in Table 2.2). Internal consistency and descriptive statistics are summarised in Table 2.3.

2.3.2.2 Incentives

Research has shown that incentives can influence the types of people who respond to surveys, which may have implications for responses (Singer & Couper, 2008). In the present study almost forty-one percent of the sample received an incentive for completing the research, in the form of credits as part of a mandatory research participation scheme within the University. While students did not have to take part in this particular study, research participation is a course requirement, and may therefore have influenced responses. It was therefore deemed important to explore if there were any significant differences between samples. Demographic information for each sample is summarised in Table 2.4. The two samples differed considerably on the age of participants, and as such this was controlled for as a possible covariate which might have also influenced responses. Two independent analyses of covariance (ANCOVAs) were conducted to explore group differences in responses for each subscale. The Levene's test indicated that homogeneity assumptions were violated, as this test is sensitive to large and unequal sample sizes differences in variance across the two groups was explored. For both the OW and AC subscales this difference was less than two (1.15 and 1.12 respectively) so it was considered possible to continue with the analyses (Field, 2013). After controlling for age, no significant effect of receiving credits for responses on either the OW subscale, $F(1, 259) = 0.18, p = .671$, or the AC subscale, $F(1, 259) = 0.58, p = .449$, emerged.

Table 2.2

Factor loadings for Exploratory Factor Analysis with Direct Oblimin Rotation for Refined Item Pool

Item	Openness & Willingness	Avoidance & Control
I allow myself to experience emotions for as long as they remain	.794	
I allow myself to experience unpleasant thoughts even if they get worse	.790	
I allow my thoughts to be there for as long as they remain	.783	
I am willing to fully experience all thoughts that come to into my mind	.770	
I allow myself to experience unpleasant emotions even if they get worse	.760	
I am willing to fully experience all emotions that arise	.741	
I allow my thoughts to come and go freely	.706	
I welcome all emotions	.706	
I allow myself to experience all of my emotions	.623	
I work hard to keep out unpleasant emotions		.867
I try to put problems out of my mind		.758
I often do things to avoid unpleasant emotions		.749
I tell myself I should not feel certain things		.731
I try to distract myself when I feel unpleasant emotions		.713
I work hard to keep out unpleasant thoughts		.688
It is important to keep emotions under control		.668
When I experience unpleasant emotions, I try to change or get rid of them		.660
I tell myself that I should not have certain thoughts		.547

Development and initial validation of the Southampton Acceptance Scale (SAS)

Table 2.3

Descriptive Statistics and Alpha Reliabilities Across Studies

Study	Sample (n)	Scale	M	SD	Minimum	Maximum	Alpha
Two	Total (268)	OW	35.36	9.94	10	63	.92
		AC	32.14	10.21	9	58	.90
	Research credits (109)	OW	33.84	8.42	16	53	.87
		AC	30.15	7.97	12	46	.83
	No incentive (159)	OW	36.44	10.78	10	63	.93
		AC	33.55	11.40	9	58	.92
Three	Total (514)	OW	38.36	9.44	14	63	.89
		AC	39.12	9.07	10	63	.84
	Young adults (246)	OW	37.95	8.79	14	62	.89
		AC	40.42	8.33	13	63	.84
	Middle aged (87)	OW	38.44	10.85	15	60	.93
		AC	37.83	8.69	20	59	.82
	Older adults (177)	OW	38.84	9.38	15	63	.87
		AC	37.83	9.97	10	61	.86

Note. OW= Openness and Willingness, AC= Avoidance and Control

Table 2.4

Demographics for Incentivised and Non-Incentivised Participants in Study Two

	Received credits (n=109)	No incentive received (n=159)
Age (years)		
M(SD)	20.22 (4.05)	32.91 (12.60)
Range	18-50	18-73
Gender, N (%)		
Female	91 (83.5)	123 (77.4)
Male	18 (16.5)	33 (20.8)
Gender Fluid/Genderqueer	0	2 (0.6)

2.3.2.3 Reading Level

Reading age of the SAS was assessed using the Flesch Reading Ease Score (FRES) and the Flesch Kincaid Grade Level. Although imperfect, especially in scales whose items are independent from one another, these tools can help give some idea as to the complexity of wording (Streiner, Norman, & Cairney, 2015). The FRES is rated on a 100-point scale, with higher scores indicative of easier reading. The SAS obtained a FRES score of 54.6, which is approximate to the reading age of thirteen.

2.4 Phase Three

There were two primary aims of Study Three. Firstly, to confirm the hypothesised factor structure from Study Two using an independent sample. Secondly, to assess the reliability and validity of the SAS. If a measure is reliable, it should operate equally in different populations and contexts. As such, the reliability of the SAS across three age groups was assessed. Furthermore, as participants were able to complete the measure either online or using pen and paper, group differences were also explored.

When developing a new measure, it is important to evidence that it measures the construct it claims to, i.e., its validity. This is often done by exploring how the new scale relates to other existing instruments, which measure either similar, related or unrelated constructs (Kline, 2000). Three aspects of validity were assessed regarding the SAS: whether it was significantly related to similar constructs (convergent validity), that it correlated with theoretically expected outcomes (concurrent and predictive validity), and that it was not strongly related to theoretically distinct constructs (discriminant validity).

Convergent validity was assessed by exploring the relationship of the SAS to the wider ACT model and psychological flexibility using the AAQ-II (Bond et al., 2011). The AAQ-II was chosen as this is a well-established ACT measure of psychological

Development and initial validation of the Southampton Acceptance Scale (SAS)

inflexibility and has often been used to measure acceptance. The relationship between the SAS and two well established emotion regulation strategies was also assessed using the Emotion Regulation Questionnaire (ERQ; Gross & John, 2003), which measures emotional suppression and cognitive reappraisal. Higher scores on the AAQ-II (indicating greater inflexibility) were expected to positively correlate with the AC subscale, while negative correlations were expected with the OW subscale. Small-to-moderate correlations were expected as the AAQ-II measures wider psychological inflexibility which includes other ACT processes in addition to acceptance/experiential avoidance. Given that emotional suppression is a form of experiential avoidance, this was also expected to correlate positively with the AC subscale. However, as the SAS assesses responses to both thoughts and emotions while the ERQ only explores emotional suppression, a small correlation was expected. A small negative correlation of suppression with OW was expected. The reappraisal subscale was not expected to correlate with either subscale on the SAS.

Concurrent validity was assessed using the Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988) and Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS; Stewart-Brown et al., 2009) as acceptance and avoidance are professed to be important determinants of psychological affect and wellbeing. OW was expected to positively correlate with positive affect (PA) and wellbeing, while negatively correlating with negative affect (NA). Conversely, the AC subscale was expected to positively correlate with NA, while negatively correlating with PA and wellbeing. Small-to-moderate correlations were expected for all variables.

The impression management (a form of social desirability) subscale of the Balanced Inventory of Desired Responding – Short form (BIDR-16; Hart, Ritchie, Hepper, & Gebauer, 2015) was also used to assess for possible desirable responding. As this is

considered a theoretically distinct concept, no association was anticipated with either subscale.

2.4.1 Method

Data for study three was collected as part of a wider study on emotion regulation by the author. Ethical approval was granted for the secondary analysis of the data for the present study by the University of Southampton Research Ethics Committee. Ethical consideration was given to the sensitive nature of the subject matter and potential negative emotional responses to taking part. Participants were provided with an information sheet which outlined confidentiality, data handling and the right to withdraw consent prior to agreeing to take part (Appendix G). After completing the study, participants were provided with a debrief sheet which also contained information about further support available should participants have experienced any negative emotions as a result of taking part (Appendix H).

2.4.1.1 Participants

Participants were community dwelling adults from a non-clinical population. The mean age of the sample was 43.61 years ($SD = 24.55$, range 18-91), and they were from United Kingdom ($n=482$, 95.4%), United States of America ($n = 10$, 1.9%), Ireland ($n=5$, 1.0%), Australia ($n=3$, 1.1%), Europe ($n=2$, 0.4%), Canada ($n = 1$, 0.2%), or not stated ($n=3$, 0.6%). The sample included students ($n= 246$, 48.0%), retired individuals ($n=174$, 33.9%), those in employment ($n=83$, 16.2%) and non-employed adults ($n=8$, 1.6%) or not stated ($n=1$, 0.2%). Unemployed individuals stated that they were stay at home parents, full-time carers, or unable to work due to medical reasons.

Almost half ($n=228$, 44.4%) of the sample identified that they had previously experienced mental health difficulties. Of these, 148 (65.3%) had received treatment for

Development and initial validation of the Southampton Acceptance Scale (SAS)

these difficulties, with 98 (43.0%) having received some psychological or talking therapy. Fifty-seven participants (25.0%) reported previous CBT, 11 (4.8%) had received mindfulness, 3 (1.3%) had attended DBT, and no participants reported having ACT.

A number of recruitment sources were used to reach adults across the lifespan. The study was advertised within the University, through posters and the online research portal. Social media was used to advertise the study (Facebook and Twitter), including on research participation and special interest groups. A database of volunteers collated by the psychology department was used to make contact with older adults, and voluntary organisations for older adults were also approached, such as the University of Third Age (U3A). Finally, an opportunity sample of personal contacts was also invited to take part.

Psychology students who were recruited through the University's research portal were awarded credits for taking part in line with the University's research participation scheme. All participants were given the opportunity to opt into a prize draw to win vouchers as a thank you for their participation.

2.4.1.2 Materials and procedure

Participants completed the newly developed 18-item acceptance questionnaire (Appendix I) in addition to the aforementioned measures to determine the validity of the SAS. These measures are described in more detail below. Demographic information was also captured.

Given the wide age range of participants, it was anticipated that some would not feel confident using, or have access to, a computer to complete the measures. To increase accessibility all participants were given the option of taking part online, or completing paper questionnaires returned using freepost envelopes.

2.4.1.2.1 The Acceptance and Action Questionnaire- Revised

The AAQ-II (Bond et al., 2011; Appendix J) is a 7-item self-report questionnaire which measures psychological flexibility and inflexibility, also known as acceptance and experiential avoidance. Items are rated using a 7-point Likert scale from 1 (*'never true'*) to 7 (*'always true'*). Scores on the AAQ-II range from 7 to 49, with lower scores indicating greater acceptance/psychological flexibility, and higher scores indicating greater levels of avoidance. The AAQ-II demonstrates good reliability (ranging from $\alpha=.78$ to $.88$), and a test-retest reliability of $.79$ (Bond et al., 2011). Internal consistency for the current study was comparable with published values ($\alpha=.85$).

2.4.1.2.2 The Emotion Regulation Questionnaire

The ERQ (Gross & John, 2003; Appendix K), is a 10-item self-report measure of two common emotion regulation strategies. The reappraisal scale measures an individual's ability to modify emotional experience through cognitive change, consisting of 6-items. The suppression scale, 4-items, measures inhibition of emotional expression behaviours. Responses are recorded on a 7-point Likert scale, from 1 (*'strongly disagree'* to 7 (*'strongly agree'*), and no items are reverse coded. Scores for reappraisal range from 6 to 42, whilst suppression scores range from 4 to 28, with higher scores indicating greater use of respective strategies. The ERQ has good internal consistency for both scales (reappraisal $\alpha = .79$, suppression $\alpha = .73$) as well as satisfactory test-retest reliability for both subscales ($.69$; Gross & John, 2003). Internal consistency for the current study was comparable with published values; reappraisal $\alpha = .85$; expressive suppression $\alpha = .76$.

2.4.1.2.3 The Positive and Negative Affect Scale

The PANAS (Watson, Clark, & Tellegen, 1988; Appendix L) is a 20-item self-report measure, consisting of two 10-item mood scales measuring positive and negative affect. Respondents rate the extent to which they have experienced particular emotions

Development and initial validation of the Southampton Acceptance Scale (SAS)

within the past two weeks, on a 5-point Likert scale from 1 (*‘very slightly or not at all’*) to 5 (*‘very much’*). The PANAS has good internal consistency when used with the general population ($\alpha=.89$ and $\alpha=.85$ for the PA and NA subscales respectively; Crawford & Henry, 2004), as well as test-retest reliability (Watson et al., 1988). Internal consistency for the current study was comparable to published examples for both subscales; PA $\alpha=.90$ and NA $\alpha=.88$.

2.4.1.2.4 The Short Warwick-Edinburgh Mental Well-being Scale

The SWEMWBS (Stewart-Brown & Janmohamed, 2008; Appendix M) is a 7-item self-report measure of psychological well-being and functioning. Items pertain to subjective positive affect and psychological functioning over the past two weeks and are rated using a 5-point Likert scale ranging from 1 (*‘none of the time’*) to 5 (*‘all of the time’*). Possible scores range from 7 to 35, with higher scores indicating greater levels of psychological well-being. SWEMWBS requires raw scores to be converted to a metric score to ensure comparability with the original 14-item scale. The metric score conversion table was used to convert raw scores in the present study in line with these recommendations (Appendix N). The shorter version of the WEMWBS was selected for use in the present study as this is superior when measuring well-being in the general population (Stewart-Brown, Tennant, Tennant, Platt, Parkinson & Weich, 2009). The scale has good test-retest reliability (.83; Stewart-Brown et al., 2009) and reliability for the present study was $\alpha=.88$.

2.4.1.2.5 The Balanced Inventory of Desired Responding - Short Form

The BIDR-16 (Hart et al., 2015; Appendix O) is a 16-item self-report scale, comprising two subscales measuring aspects of socially desirable responding (SDR); impression management (IM, conscious biases in responses to create a socially desirable image) and self-deceptive enhancement (SDE), a non-conscious bias to perceive oneself

favourably resulting in honest but positively biased responses). Participants respond to a series of items on a 7-point Likert scale from 1 (*‘strongly disagree’*) to 7 (*‘strongly agree’*). The scale has good internal consistency for both subscales (ranging from $\alpha = .63$ to $\alpha = .82$; Hart et al, 2015). Only the IM subscale only was used in the present study to assess conscious attempts to provide SDR. This had acceptable internal consistency in the present study, $\alpha = .72$.

2.4.1.3 Analytic Strategy

2.4.1.3.1 Confirmatory factor analyses

CFA is a technique that is used once the researcher has some knowledge and understanding of latent variables and their relationship to observed variables (Byrne, 2010). Expected relationships are defined a priori in a model, and the hypothesised structure is then statistically tested. In scale development, it is common practice to follow up EFA with a CFA model to determine whether the hypothesised factor structure is retained.

Three analytic strategies were employed with the aim of developing and confirming a theoretically consistent model (Jöreskog, 1993). Firstly, a purely confirmatory approach was used to test the fit of the 18-item two factor model proposed in Study Two. Secondly, a model generating strategy was used with the aim of preserving the two-factor structure, whilst improving model fit through refining the model. Thirdly, an alternative model strategy was used to test whether the two-factor solution was a better fit than a single factor structure.

Next, as we sought to develop a measure that could be used with a range of populations, we assessed whether the 18-items of the SAS functioned in a similar way across adults of different age groups. Participants were grouped into three age groups, young (18-30 years), middle (31-64) and older adults (65 and over).

Development and initial validation of the Southampton Acceptance Scale (SAS)

Goodness of fit for each CFA model was assessed using Maximum Likelihood chi-square (χ^2), comparative fit index (CFI), and root square mean error of approximation (RMSEA). A good-fitting model is indicated by a non-significant chi-square test, CFI indices of at least .90 (Bentler & Bonett, 1980), and RMSEA indices below .80 (Bentler, 1990). It has previously been suggested that the model chi-square test is overly conservative and sensitive to sample size, thus a non-significant chi-square test was unlikely in the present study.

2.4.1.3.2 Construct Validity

A series of bivariate correlations were used to examine relationships between theoretically related and unrelated constructs. Given the number of comparisons, false discovery rate (FDR) analyses were undertaken to control for family-wise error rates. FDR was selected over traditional Bonferroni corrections as FDR-based comparisons are more powerful and values have a monotonic linear increase, as opposed to the Bonferroni which has a fixed, and often overly conservative, threshold (Pike, 2011).

2.4.1.3.3 Format invariance

As participants were able to choose the method of completion, t-tests were conducted to determine whether there were any differences in responses for those who completed the online and paper version of the measures.

2.4.2 Results and discussion

2.4.2.1 Confirmatory Factor Analyses

Data exploration indicated missing data, which was less than 1% and missing at random. Full Information Maximum Likelihood (FIML) estimations can be used within AMOS to cope with missing data rather than listwise or pairwise deletion. FIML estimations are considered to be superior to deletion approaches to missing data as they

reduce the risk of bias, preserve power and have minimal impact on model fit (Byrne, 2010). As such, FIML imputations was used throughout the following analyses.

2.4.2.1.1 A priori measurement model

The 18-item, two-factor solution was subject to a CFA using AMOS 25 (Arbuckle, 2017). Goodness of fit indicators suggested a poor fit to the data: $\chi^2(134, N= 513) = 822.18, p < .001$ CFI = .83, RMSEA = .10 (90% confidence interval) [CI] = [.094, .107]. A model generating strategy was therefore adopted to refine the SAS, and a series of two-factor models were tested.

2.4.2.1.2 Model generating

Observation of the modification indices indicated that all indicators loaded significantly and in the expected direction onto the specified latent factors, with no cross loadings. As expected, the two latent factors demonstrated a small, negative covariance, indicating that they were indeed distinct constructs (-.35). However, the modification indices suggested misspecified error covariances, where covariation was underrepresented by the latent factor structure. Exploration of items and recommended respecifications indicated that this was likely to have occurred due to a high level of content overlap in items (Byrne, 2010). Consequently, the model was respecified with covariances between identified error terms where this was theoretically coherent. The model was tested for comparable fit after each respecification in line with best practice, and these stages are summarised in Table 2.5. While a non-significant chi-square difference was not obtained, no further covariances were added beyond model seven as there was no theoretical justifications for respecifying the model. The final model is represented in Figure 2.1.

2.4.2.1.3 Alternative model: Assessing dimensionality

An 18-item, one-factor model was tested, which demonstrated a poorer fit to the data than the initial two-factor model: $\chi^2(135, N= 513) = 1826.82, p < .001$ CFI = .59, RMSEA = .156 (90% confidence interval [CI] = [.150, .163]). Chi-square difference between the models was 1004.64, with 1 df difference. Given this value is in excess of 1000, it was not possible to compute a p value; however, the Akaike information criterion (AIC) values of the one- and two-factor models (1934.82 and 932.18, respectively) demonstrated that the two-factor solution was a better fit to the data as the AIC value was smaller (Tabachnick & Fidell, 2014).

Table 2.5

Model comparisons for respecification of CFA model

Model	χ^2	df_M	$\Delta\chi^2$	CFI	RMSEA [CI]
0. Baseline	822.18	134	--	.83	.100 [.094, .107]
1. Covariance e4-e7	673.64	133	148.54**	.87	.089 [.083, .096]
2. Covariance e17-e5	582.55	132	91.09**	.89	.082 [.075, .089]
3. Covariance e14-e15	511.26	131	162.38**	.91	.075 [.069, .082]
4. Covariance e2-e3	441.06	130	70.20**	.92	.068 [.061, .075]
5. Covariance e17-e9	402.40	129	38.66**	.93	.064 [.057, .072]
6. Covariance e3-e9	380.39	128	22.01**	.94	.062 [.055, .069]
7. Covariance e2-e4	364.93	127	15.461**	.94	.061 [.053, .069]

Note. **significant at $p < .001$

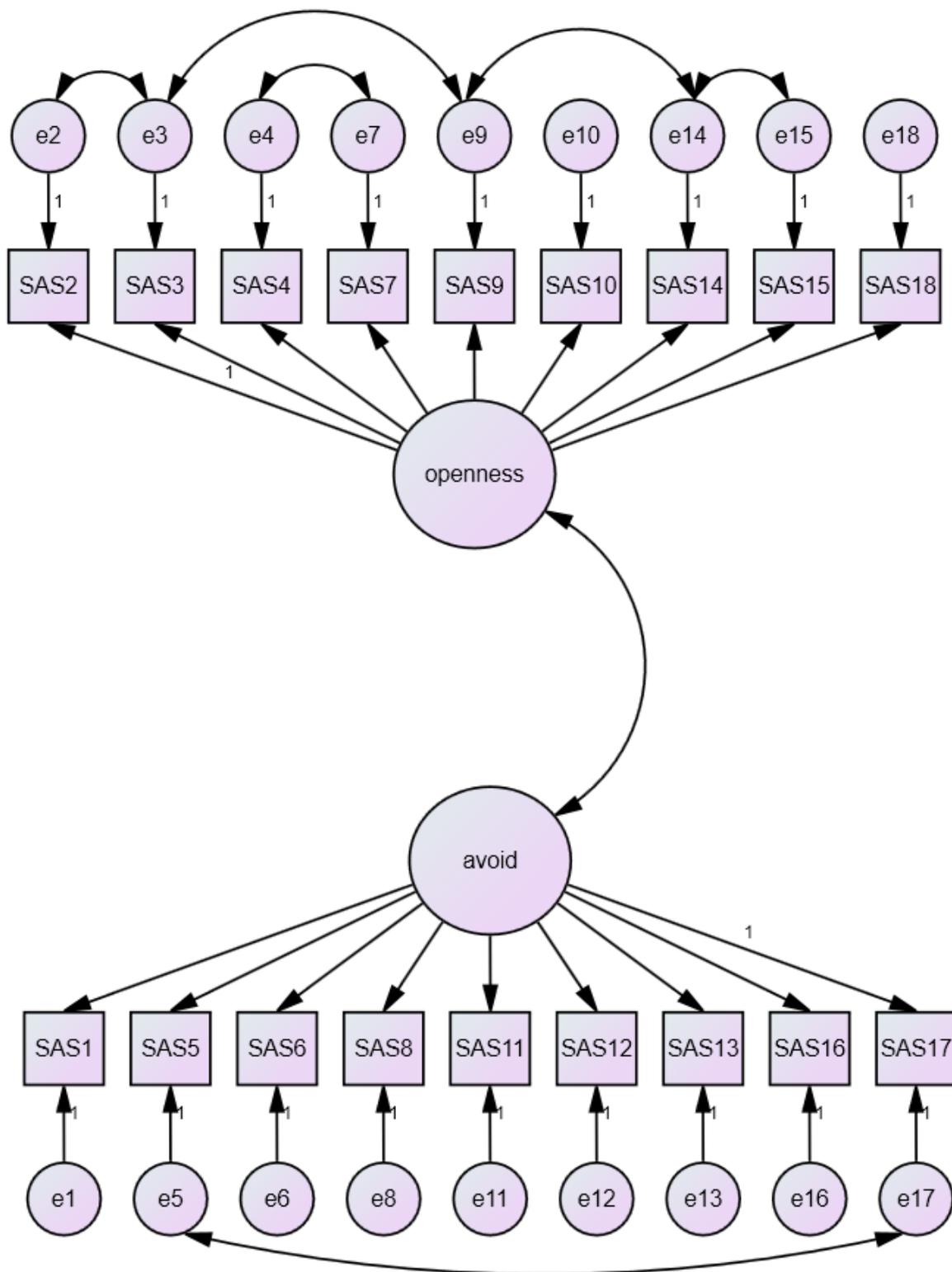


Figure 2.1 Respecified model of factorial structure for the Southampton Acceptance Scale (SAS) Model 7.

2.4.2.1.4 Measurement Invariance: Age

As a well-fitting measurement model had been identified for the pooled data set, it was possible to proceed and determine whether individual items performed differently as a function of age, using a multigroup approach. The sample was divided into three age groups: younger adults ($n= 246$), middle aged adults ($n= 81$), and older adults ($n= 177$). Testing for invariance across groups is conducted in multiple stages, whereby a baseline model of the two groups in which all parameters are free to vary (i.e., take on different values for each group), known as the configural model, is first estimated. Constraints are then systematically added to different parts of the model to test for invariance at each level, and these nested models are compared to the configural model to determine whether model fit significantly decreases using the chi-square difference statistic ($\Delta \chi^2$; Byrne, 2010).

As the residual error covariances were important to overall model fit for all groups, these were treated as equal across groups for all analyses, including the configural model in line with recommendations by Byrne (2010). The configural model, in which all groups are estimated simultaneously, was shown to be a good fit to the data $\chi^2(df= 396) = 783.83$, $p < .001$; CFI = .908; RMSEA .044, C.I.[.040, .049], $p < .001$.

Next a measurement weights model was tested, in which all unstandardised factor loadings were constrained equal across both age groups. This model had no significant impact on model fit, demonstrating invariance across groups (i.e. items of the SAS designed to measure acceptance and avoidance operate equally across age groups), as demonstrated by a non-significant chi-square difference statistic, see Table 2.6.

The structural aspect of the model was then tested for invariance. In the structural covariances model, the covariance between the two identified factors were constrained to be equal, which did not detriment model fit, suggesting that covariance between

acceptance and avoidance was invariant across groups. These results are summarised in Table 2.6.

Table 2.6
Goodness of Fit Statistics for Tests of Multigroup Invariance Across Age Groups

Constraints	χ^2 (df)	<i>p</i>	$\Delta \chi^2$ (df)	<i>p</i>	CFI	RMSEA [C.I.]
1. Configural Model	783.83 (396)	**	--	--	.908	.044 [.040, .049]
2. Measurement Weights	810.63 (428)	**	5.2864 (11)	.728	.909	.042 [.038, .047]
3. Structural covariances	821.47 (434)	**	53.5656 (33)	.486	.908	.042 [.038, .047]

Note. ** $p < .00001$

2.4.2.2 Validity of the SAS

Table 2.7 summarises the associations between the SAS and theoretically similar and distinct variables in the present study. Results are reported with FDR adjusted significant levels.

2.4.2.2.1 Convergent validity

As expected the AC subscale demonstrated small-to-moderate positive correlations with the AAQ-II and the suppression subscale of the ERQ. Unexpectedly, the AC also demonstrated a small significant correlation with the ERQ reappraisal subscale.

Reappraisal is defined as a form of cognitive change which seeks to alter the emotion elicited (Gross & John, 2003), which can be either in terms of process (*‘When I want to feel more positive emotion (such as joy or amusement), I change the way I’m thinking*

Development and initial validation of the Southampton Acceptance Scale (SAS)

about the situation') or content ('*When I want to feel more positive emotion I change what I'm thinking about*'). Consequently, there is likely to be a degree of overlap with items on the AC subscale that pertain to processes such as distraction, as these involve changing the focus of thoughts (e.g. '*I try to put problems out of my mind*').

In line with our hypotheses, the OW subscale was negatively correlated with both the AAQ-II and ERQ suppression, while no relationship was found with reappraisal.

2.4.2.2.2 Concurrent validity

As expected, the OW scale demonstrated significant small positive correlations with both PA and wellbeing. Similarly, AC and NA were also significantly positively correlated. While OW was negative correlated with NA, and AC was negatively correlated with PA and wellbeing, these relationships were less strong. Correlations are summarised in Table 2.7.

2.4.2.2.3 Discriminant validity

As expected, no association was found between the SAS subscales and social desirability.

2.4.2.3 Format invariance

No younger or middle-aged adults requested to complete the study using paper measures. In the older adult sample, 83 participants, of 223 (37.22%), requested paper packs of the measures. A t-test indicated no significant differences between older adults who completed the SAS online or in paper format for either the OW $t(221) = 0.17, p = .866$, or the AC subscale $t(221) = 0.11, p = .911$.

Table 2.7

Correlations between the SAS Subscales with Existing Scales in Study Three

Measure	Southampton Acceptance Scale				AAQ-II	
	Openness and Willingness		Avoidance and Control		<i>r</i>	<i>R</i> ²
	<i>r</i>	<i>R</i> ²	<i>r</i>	<i>R</i> ²		
AAQ-II	-.25**	.062	.41**	.170	--	--
ERQ						
Reappraisal	-.02	.004	.21**	.044	-.18**	.032
Suppression	-.25**	.062	.22**	.048	.18**	.032
PANAS						
Positive Affect	.26**	.065	-.10*	.010	.48**	.230
Negative Affect	-.19**	.037	.34**	.114	.61**	.372
SWEMWBS	.34**	.114	-.19**	.038	.61**	.372
BIDR-IM	.02	.004	.06	.004	.002	.000

Note. AAQ-II= Acceptance and Action Questionnaire-Revised; ERQ= Emotion Regulation Questionnaire; SWEMWBS= Short Warwick-Edinburgh Mental Wellbeing Scale; BIDR-IM= Balanced Inventory of Desired Responding Short Form- Impression Management Subscale. All significance levels are FDR adjusted, **significant at $p < .001$, *significant at $p < .05$.

2.5 General Discussion

There are presently no established measures of acceptance which assess the construct as a general process as defined within the ACT model. Existing ACT measures which assess acceptance either do so with regards to acceptance of a particular difficulty, or in the context of wider conceptualisations of psychological flexibility. Furthermore, concerns

Development and initial validation of the Southampton Acceptance Scale (SAS)

have been raised regarding the validity of extant measures. As such, the present study sought to develop and validate a new measure of acceptance to meet this need.

2.5.1 Dimensionality

The measure development process resulted in an 18-item, two factor scale. One factor contained items relating to acceptance processes, while the other contained experiential avoidance items. This factor structure was then explored using a large sample of adults from across the lifespan. In contrast to many published scales, the SAS does not contain any reverse-scored items; these were excluded following factor analyses due to poor loadings. Some research indicates that positive and negatively worded items are likely to load onto distinct factors due to method effects, rather than to substantial differences in construct validity (DiStefano & Motl, 2006). However, in the present study, during the initial EFA when no factors were specified, this was not the case. Items of both positive and negative-valence loaded onto the same factor, although they demonstrated differential factor loading strengths. This replicates Gamez et al.'s (2011) findings which showed that reverse-scored acceptance items did not satisfactorily load onto the non-acceptance scale. As such, while method effects may be present, the findings of the present study indicate that there may also be important differences in the latent structure that these items are measuring. As the SAS is the only scale of acceptance which currently distinguishes avoidance and acceptance and measures them on separate scales, rather than relying on inverse scoring, further work is needed to fully understand this relationship.

Theoretically, both acceptance and experiential avoidance are considered as active processes; acceptance describes the process of actively moving towards private experiences with openness and curiosity, whereas experiential avoidance is the process of trying to avoid, control or escape private events. As such, the absence of one such approach to private experiences would not automatically result in the other; i.e., it is

theoretically possible to be low in avoidance without being high in acceptance.

Consequently, acceptance and experiential avoidance may represent distinct, albeit related constructs, much like positive and negative affect. Since this reconceptualization of affect, models recognised the need to address both the down-regulation of negative affect and up-regulation of positive emotions. While some strategies can facilitate both processes, this is not always the case. Consequently, if experiential avoidance and acceptance are indeed distinct constructs, it will be important for clinical research to explore whether strategies which seek to foster acceptance also serve to reduce avoidance, or if further strategies are required that explicitly address avoidance.

2.5.2 Reliability and validity

The SAS had excellent internal reliability across samples. The SAS demonstrated invariance with adults at different life stages (young, middle-aged, and older adults) despite unequal sample sizes, suggesting that scale items are not susceptible to cohort effects. Predicted patterns of relationships with theoretically related and distinct constructs were also found.

As expected, the SAS subscales demonstrated small to moderate correlations with AAQ-II. The AC subscale demonstrated a stronger positive correlation with the AAQ-II, suggesting that experiential avoidance is a significant component of psychological inflexibility, as measured by the AAQ-II. The acceptance subscale demonstrated a small negative correlation, which again raises queries regarding the dimensionality of acceptance and avoidance.

In the present study, acceptance was not associated with reappraisal, however, avoidance was. Many items on the AC scale refer to change agendas which could be considered synonymous with the cognitive change advocated in reappraisal, and may explain some of the shared variance. Both acceptance and reappraisal require individuals to

Development and initial validation of the Southampton Acceptance Scale (SAS)

step back from and observe private events. However, the response to these differs (allowing versus changing), with both approaches having clinical merit. It has been suggested that it is the act of ‘decentering’ before responding to events that is significant, and that both acceptance and reappraisal require mindful awareness of events prior to responding (Teasdale, 1999). The lack of association between the two measures would suggest that the scales distinguish responses to private events from the ‘decentering’ process. Other published ACT measures have not directly compared acceptance and reappraisal, so further work is needed to understand this relationship.

Both acceptance and avoidance were associated with emotionally expressive suppression. Small correlations were anticipated as the measure focused on a specific type of suppression: emotional expression. Furthermore, suppression is one of many possible experiential responses. The MEAQ (Gámez et al., 2011) identified six types of experiential avoidance, one of which was distraction/suppression. Further research is needed to understand how these different avoidance processes relate to the SAS.

The ACT model suggests that symptom reduction is not the aim of acceptance, rather it is living in line with values alongside private experiences, although many individuals do report some symptom reduction. Despite affect modulation not being the aim of acceptance, relationships between affect and acceptance were found. Acceptance demonstrated a stronger relationship with psychological wellbeing than with affect. Given acceptance considers elements pertinent to valued living, this stronger relationship is perhaps unsurprising.

Correlations of the SAS scales with all variables were smaller than those found in published work for many other ACT process measures. There are several possible reasons for this. Firstly, the SAS seeks to measure two specific processes within the ACT model, whereas measures such as the AAQ-II and CompACT try to measure the model in its

entirety. Consequently, it is unsurprising that the proportion of variance explained is smaller. While some individual process measures, such as the CFQ, demonstrate comparable relationships with affect scales as general measures, it has been proposed that this may result from interdeterminance of the constructs under consideration (Gillanders et al., 2014). Other process measures, such as the Valued Living Questionnaire (VLQ; Wilson, Sandoz, Kitchens, & Roberts, 2010) demonstrate comparable relationships with two of the ACT processes - affect and wellbeing - as does the SAS. Secondly, scales such as the CFQ were developed using items from existing measures, resulting in shared items which artificially increases shared variance and strengthens the relationships between constructs.

Age may also have influenced the strength of relationships between variables in the present study. The sample in study three contained substantially more older adults than are commonly used in such studies. For example, the VLQ was developed and validated using a sample where 93.2% of adults were aged 18-22. The SAS demonstrated measurement invariance and so items can be considered to operate equally across the lifespan. However, relationships between emotion regulation strategies and affect have been shown to differ as a function of age (Allen & Windsor, 2017). As such, the strength of relationships between the SAS and other scales may be a result of age-related differences in these constructs.

2.5.3 Strengths

There are a number of strengths in the present study. Firstly, both experts and non-experts were involved in reviewing items from the initial pool supporting the construct validity of this measures. Many existing scales rely solely on experts for item selection, despite non-expert involvement being an essential criterion for determining content validity. Furthermore, many scales are developed within specific populations and may therefore lack generalisability. A strength of this study is that adults from across the life

Development and initial validation of the Southampton Acceptance Scale (SAS)

span were involved in all phases of scale development, and the SAS was invariant across age groups. As such, the SAS is suitable for use with adults across the lifespan. The study also explored whether completion method affected responses, and no differences were found between those who completed the measure online or those who used paper copies.

2.5.4 Limitations

In addition to the strengths of the present study, there are a number of methodological limitations which should be taken into consideration. Firstly, with regards to sample limitations, across all studies the samples had a majority of female participants. As this was an initial development study, all samples were adults from a non-clinical population. While 45% of the sample identified themselves as having experienced a mental health difficulty in the past, it is not possible to draw conclusions as to how the SAS would perform in a clinical population. A number of participants had previously received CBT or psychological therapies that advocate acceptance. However, these samples were too small to compare whether there were group differences in scores on the SAS. Furthermore, while the present study demonstrated that the SAS was invariant at both the item and factor level across age groups, its performance in across other populations is yet to be tested. A further limitation is that the responsiveness of the SAS to treatment has not yet been explored. A limitation of general ACT process measures is that it is not possible to determine specific mechanisms of change throughout treatment.

2.5.5 Directions for future research

To build upon the initial validation of the SAS, a number of further studies would be beneficial to explore how the measure performs in different settings. Firstly, future research should seek to test the factor structure and explore the validity of the SAS using a

clinical sample, to determine whether it performs comparably to the non-clinical samples used here.

The present study explored relationships of the SAS with extant measures related to psychological flexibility and emotion regulation strategies. Although these results show preliminary evidence for the validity of the SAS, its relationship to other key ACT processes is yet to be investigated. Psychological inflexibility, as measured by the AAQ-II, accounted for more variance in scores of affect and wellbeing, which is to be expected. However, further research is required to determine whether measures of individual ACT processes can cumulatively explain variance in psychological wellbeing. Furthermore, given that acceptance and experiential avoidance are theorised to predict committed action, which in turn results in psychological wellbeing/distress, future studies should seek to test a predictive model of these processes.

As the responsiveness of the SAS is yet to be explored, it will be important for future research to determine whether the SAS is sensitive to treatment effects so that this shortcoming can be addressed. Once this has been established, the SAS can be used alongside other specific process measures to test hypotheses around mechanisms of change and efficacy of acceptance-based interventions.

The present study sought to further explore the dimensionality of acceptance and experiential avoidance. While the results suggest that these are two distinct constructs, there is not enough evidence to demonstrate that this finding is free of method effects. The SAS is unique in that it is the only acceptance scale that does not rely on reverse scored items. Further research is therefore needed to explore the dimensionality and impact of reverse scoring and negatively valenced items of extant measures.

2.5.6 Clinical implications and potential uses

The results of the present study demonstrate the value of psychometrics in relation to clinical models. The need for valid and reliable measures to assess latent constructs is imperative for evaluating treatments which seek to address these. Prior to the development of the SAS, there was no acceptance measure consistent with the ACT model. The different subscales identified during the development of the SAS allow researchers and clinicians to measure both acceptance and experiential avoidance separately, without relying on reverse scoring items. This is likely to be particularly important given the uncertainty regarding the dimensionality of these two constructs.

As the measure does not require any previous knowledge or understanding of ACT and its constructs, it is suitable for use with a wide range of non-clinical populations. This, in addition to its simplicity and brevity, make the SAS an ideal screening tool for both clinical and research settings.

2.5.7 Conclusions

Although in its infancy, the SAS is a theoretically coherent measure which demonstrates strong psychometric properties. The brevity and simplicity of the SAS mean that it is well suited for use in clinical and research settings for multiple purposes. Furthermore, as the measure is generic to all private events, it is applicable to a wide range of populations and contexts. While the present study demonstrates that the SAS shows promise as a measure of acceptance, further research is needed to empirically determine its robustness across other populations, further explore psychometric properties and examine its clinical utility.

Appendix A COSMIN manual summary table for definitions of domains, measurement properties, and aspects of measurement properties

Term			Definition
Domain	Measurement property	Aspect of a measurement property	
Reliability			The degree to which the measurement is free from measurement error
Reliability (extended definition)			The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same health related-patient reported outcomes (HR- PRO) (internal consistency); over time (test-retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e. raters or responders) on different occasions (intra-rater)
	Internal consistency		The degree of the interrelatedness among the items
	Reliability		The proportion of the total variance in the measurements which is due to 'true' [†] differences between patients
	Measurement error		The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
Validity			The degree to which an HR-PRO instrument measures the construct(s) it purports to measure
	Content validity		The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured
		Face validity	The degree to which (the items of) an HR-PRO instrument indeed looks as though they are an adequate reflection of the construct to be measured
	Construct validity		The degree to which the scores of an HR-PRO instrument are consistent with hypotheses (<i>for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups</i>) based on the assumption that the HR-PRO instrument validly measures the construct to be measured
		Structural validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured
		Hypotheses testing	Idem construct validity

Appendix A

		Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
	Criterion validity		The degree to which the scores of an HR-PRO instrument are an adequate reflection of a 'gold standard'
Responsiveness			The ability of an HR-PRO instrument to detect change over time in the construct to be measured
	Responsiveness		Idem responsiveness
Interpretability*			Interpretability is the degree to which one can assign qualitative meaning - that is, clinical or commonly understood connotations – to an instrument's quantitative scores or change in scores.

Appendix B Quality criteria for measurement properties of health status questionnaires (Terwee et al., 2007)

Property	Definition	Quality Criteria
1. Content Validity	The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire	<p>+ A clear description is provided of the measurement aim, the target population, the concepts that are being measured, and the item selection AND target population and (investigators OR experts) were involved in item selection;</p> <p>? A clear description of above-mentioned aspects is lacking OR only target population involved OR doubtful design or method;</p> <p>- No target population involvement;</p> <p>0 No information found on target population involvement.</p>
2. Internal Consistency	The extent to which items in a (sub)scale are intercorrelated, thus measuring the same construct	<p>+ Factor analyses performed on adequate sample size ($7 * \# \text{ items}$ and ≥ 100) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95;</p> <p>? No factor analysis OR doubtful design or method;</p> <p>- Cronbach's alpha(s) < 0.70 or > 0.95, despite adequate design and method;</p> <p>0 No information found on internal consistency.</p>
3. Criterion Validity	The extent to which	+ Convincing arguments that gold

scores on a particular questionnaire relate to a gold standard

standard is “gold” AND correlation with gold standard ≥ 0.70 ;

? No convincing arguments that gold standard is “gold” OR doubtful design or method;

- Correlation with gold standard < 0.70 , despite adequate design and method;

0 No information found on criterion validity.

4. Construct Validity

The extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured

+ Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses;

? Doubtful design or method (e.g., no hypotheses);

- Less than 75% of hypotheses were confirmed, despite adequate design and methods;

0 No information found on construct validity.

5. Reproducibility

5.1. Agreement

The extent to which the scores on repeated measures are close to each other (absolute measurement error)

+ MIC $<$ SDC OR MIC outside the LOA OR convincing arguments that agreement is acceptable;

? Doubtful design or method OR (MIC not defined AND no convincing arguments that agreement is acceptable);

- MIC \geq SDC OR MIC equals or inside LOA, despite adequate design and method;

0 No information found on agreement.

5.2. Reliability

The extent to which + ICC or weighted Kappa ≥ 0.70 ;

	patients can be distinguished from each other, despite measurement errors (relative measurement error)	? Doubtful design or method (e.g., time interval not mentioned); - ICC or weighted Kappa < 0.70, despite adequate design and method; 0 No information found on reliability
6. Responsiveness	The ability of a questionnaire to detect clinically important changes over time	+ SDC or SDC < MIC OR MIC outside the LOA OR RR > 1.96 OR AUC ≥ 0.70; ? Doubtful design or method; - SDC or SDC ≥ MIC OR MIC equals or inside LOA OR RR ≤ 1.96 OR AUC < 0.70, despite adequate design and methods; 0 No information found on responsiveness.
7. Floor and Ceiling Effects	The number of respondents who achieved the lowest or highest possible score	+ ≤15% of the respondents achieved the highest or lowest possible scores; ? Doubtful design or method; - >15% of the respondents achieved the highest or lowest possible scores, despite adequate design and methods; 0 No information found on interpretation.
8. Interpretability	The degree to which one can assign qualitative meaning to quantitative scores	+ Mean and SD scores presented of at least four relevant subgroups of patients and MIC defined; ? Doubtful design or method OR less than four subgroups OR no MIC defined; 0 No information found on interpretation.

Appendix C ACT Experts Information Sheet

I am a trainee Clinical Psychologist at the University of Southampton. My doctoral thesis is investigating the influence of ageing and resilience on emotion regulation and affect. As part of this project we will be investigating the use of different emotion regulation strategies and implications for wellbeing. Acceptance is one of the strategies of interest

Current measures of acceptance have been criticised for being unable to discriminate process and outcome, as they also assess valued living and distress (Chawla & Ostafin, 2007; Wolgast, 2014). Consequently, we are developing a new measure of acceptance within the context of emotion regulation; focussing on the process of acceptance, rather than the outcome (effective living and distress). As part of this process, we are inviting clinicians and researchers with expertise in Acceptance and Commitment Therapy (ACT) to provide feedback on a series of potential items that will form the basis of this scale.

Why have I been chosen?

Clinicians and researchers with expertise in ACT are invited to give their professional opinion on the item pool.

Do I have to take part?

It is completely your choice whether you wish to take part. You are free to withdraw your consent to participate in the study at any time, without penalty.

What are the benefits of taking part?

By taking part in this research, you will be contributing to the development of a new measurement tool.

What does participation involve?

We ask that you review the 60 items, considering how closely this item relates to the concept of acceptance. There is also space to comment on how clear and concise questions are, give feedback on wording, structure or comment on anything that stands out to you about the item. There is also an opportunity at the end to suggest any other items, should you wish to. It is anticipated that this will take 20 minutes to complete.

What happens to the information I give?

Any information that you give will be kept strictly confidential, and will not be released to, or viewed by, anyone other than researchers involved in this study. All data will be stored

in an anonymous format, and results of this study will not include any identifying characteristics.

Who has reviewed the study?

This study has been reviewed by the School of Psychology Research Ethics Committee, University of Southampton.

Contact for further information

If you would like more information about any aspect of the study, please do not hesitate to contact Zoe McAndrews, email: zm1g14@soton.ac.uk at School of Psychology, Building 44A, Highfield Campus, University of Southampton, SO17 1BJ.

Finally, if you have any questions about your rights as a participant in this research, please contact the Chair of the Ethics Committee, School of Psychology, University of Southampton, Southampton, SO17 1BJ, Tel: 023 80593995 or email fshs-rso@soton.ac.uk

Thank you for taking the time to read this information.

Researchers:

Zoe McAndrews, Trainee Clinical Psychologist, University of Southampton,
zm1g14@soton.ac.uk

Dr Lusia Stopa, Professor of Clinical Psychology, University of Southampton,
L.Stopa@soton.ac.uk

Dr Claire Hart, Lecturer in Psychology, University of Southampton,
C.M.Hart@soton.ac.uk

Dr Laura Brummer, Principal Clinical Psychologist, Southern Health NHS Foundation Trust, Laura.brummer@nhs.net

Ethics ID: 25199

References:

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. DOI: 10.1002/jclp.20400

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

Appendix D Ethical Approval

Approved by Faculty Ethics Committee - ERGO II 25199.A8

The logo for the University of Southampton, featuring the text "UNIVERSITY OF" in a small, uppercase font above the word "Southampton" in a larger, bold, serif font. The logo is white and is set against a dark teal background.

ERGO II – Ethics and Research Governance Online <https://www.ergo2.soton.ac.uk>

Submission ID: 25199.A8

Submission Title: Emotion Regulation and Wellbeing across the
Lifespan: The Role of Life Stressors and Resilience (Amendment 8)

Submitter Name: Zoe McAndrews

Your submission has now been approved by the Faculty Ethics
Committee. You can begin your research unless you are still awaiting
any other reviews or conditions of your approval.

Appendix E Emotion Regulation and Wellbeing across the Lifespan: The Role of Life Stressors and Resilience

Part One - Developing a new Questionnaire

Information Sheet

You are being invited to take part in a research study. Before you decide whether or not you wish to take part, it is important that you to understand the purpose of the research and what it involves. Please take some time to read all of the following information carefully. Do not hesitate to ask us if there is anything that is not clear or if you would like more information. You can also discuss it with others if you wish. Take time to decide whether or not you wish to take part. Thank you for reading this.

What is the purpose of this study?

As part of our research exploring how ageing and life events affect the way that we respond to our emotions and emotional situations, we are developing a new questionnaire to explore ways that individuals manage their emotions.

Why have I been chosen?

Anyone over the age of 18 years-old is eligible to take part in the study, and we hope to recruit around 600 people.

Do I have to take part?

It is completely your choice whether you wish to take part. If you choose to complete the questionnaire on line, please tick the box below to confirm your consent to take part. You are free to withdraw your consent to participate in the study at any time, without penalty.

What will happen if I decide to take part?

If you decide to take part you will be asked to complete a brief sheet that describes you (such as your age and gender) along with a further questionnaire. This should take approximately 5-10 minutes.

Are there any benefits in my taking part?

By taking part in this study, you will be helping us to develop a new questionnaire that looks at how people manage their emotions. It is hoped that this questionnaire will support further research into emotion regulation.

What if I change my mind?

If at any time you change your mind, and no longer wish to take part, you can withdraw your consent without any negative consequences.

What happens to the information I give?

Any information that you give will be kept strictly confidential, and will not be released to, or viewed by, anyone other than researchers involved in this study. All data will be stored in an anonymous format, and results of this study will not include your name or any other identifying characteristics.

What will happen to the results of the research?

The results of the study will be written up for publication in a psychological journal. A brief summary of the findings can be made available upon request.

Who has reviewed the study?

This study has been reviewed by the School of Psychology Research Ethics Committee, University of Southampton.

Contact for further information

If you would like more information about any aspect of the study, please do not hesitate to contact Zoe McAndrews by emailing: zm1g14@soton.ac.uk or writing to School of Psychology, Building 44A, Highfield Campus, University of Southampton, SO17 1BJ.

Finally, if you have any questions about your rights as a participant in this research, please contact the Chair of the Ethics Committee, School of Psychology, University of Southampton, Southampton, SO17 1BJ, Tel: 023 80593995 or email fshs-rso@soton.ac.uk

Thank you for taking the time to read this information.

Zoe McAndrews, Trainee Clinical Psychologist, University of Southampton

Dr Lusía Stopa, Professor of Clinical Psychology, University of Southampton

Dr Claire Hart Lecturer in Personality/Social Psychology, University of Southampton

Dr Laura Brummer, Principal Clinical Psychologist, Southern Health NHS Foundation

Trust

Appendix F Emotion Regulation and Wellbeing across the Lifespan: The Role of Life Stressors and Resilience

Part One - Developing a new Questionnaire.

Debriefing statement

Background

Research suggests that as we grow older and have different experiences, the way in which we manage our emotions may change. To support us to study these difference further, this study aimed to develop a new questionnaire that explores different ways that individuals manage their emotions.

Methodology

Participants were asked to complete questionnaires from a pool of items, which aim to explore how people tend to respond to emotional events that may be stressful or difficult. Descriptive information relating to characteristics, such as age and gender, were also collected in order to describe the participants who took part.

Results

The results will be written up and submitted for publication in a peer reviewed journal. This article will summarise the overall findings of the study, but will not identify any individuals. A brief summary of the findings can be made available to participants upon request.

Your responses

If anything that you have been asked to complete as part of this study has caused you to experience any negative emotions with which you are having difficulty coping, there are a number of sources of support available to you:

1. You can contact your general practitioner (GP).

2. If you are registered with a Southampton GP, you can contact the Steps2Wellbeing IAPT service, at:

Third Floor, Grenville House, Nelson Gate, Southampton, SO15 1GX

Tel. 0800 612 7000 or 02380 272000

Email: sstw@dhuft.nhs.uk

Web: www.steps2wellbeing.co.uk

If you do not live in Southampton, you can find information about your local Improving Access to Psychological Therapies (IAPT) service here:

[http://www.nhs.uk/Service-Search/Psychological%20therapies%20\(IAPT\)/LocationSearch/10008](http://www.nhs.uk/Service-Search/Psychological%20therapies%20(IAPT)/LocationSearch/10008)

Finally, if you have any questions about your rights as a participant in this research, or you feel that you have been placed at risk, please contact the Chair of the Ethics Committee, School of Psychology, University of Southampton, Southampton, SO17 1BJ, Tel: 023 80593995 or email fshs-rso@soton.ac.uk.

Thank you for taking the time to read this information.

Zoe McAndrews, Trainee Clinical Psychologist, University of Southampton

Dr Lusia Stopa, Professor of Clinical Psychology, University of Southampton

Dr Claire Hart Lecturer in Personality/Social Psychology, University of Southampton

Dr Laura Brummer, Principal Clinical Psychologist, Southern Health NHS Foundation Trust

More studies are available on: <http://psych.hanover.edu/research/exponnet.html> and

www.onlinepsychresearch.co.uk

Appendix G Attitudes and Emotions Across the Lifespan: Information Sheet

You are being invited to take part in a research study. Before you decide whether or not you wish to take part, it is important that you to understand the purpose of the research and what it involves. Please take some time to read all of the following information carefully. Do not hesitate to ask us if there is anything that is not clear or if you would like more information. You can also discuss it with others if you wish. Take time to decide whether or not you wish to take part. Thank you for reading this.

What is the purpose of this study?

Research suggests that as we get older and have different experiences, the way in which we manage our emotions may change. This study aims to explore how ageing and different life events affect the way that we respond to our emotions and emotional situations.

Why have I been chosen?

You have been asked if you would like to take part as you are aged either between the ages of 18-30, 40-55 or over 65 years of age. Anyone in these age groups is eligible to take part in the study, and we hope to recruit around 600 people.

Do I have to take part?

It is completely your choice whether you wish to take part. If you return a completed copy of the questionnaires, or if you complete these questionnaires on line, then we will take this as you providing informed consent to participate in this study. You are free to withdraw your consent to participate in the study at any time, without penalty.

What will happen if I decide to take part?

If you decide to take part you will be asked to complete a brief sheet that describes you (such as your age and gender) along with a series of questionnaires that ask about your emotions and how you cope. This should take approximately 25-30 minutes. You do not

have to complete these in a single sitting if this is not possible, but we ask that you complete all questionnaires within the space of one week.

Are there any benefits in my taking part?

By taking part in this study, you will be contributing to our knowledge of how people manage their emotions throughout their life. We hope that this information will inform the development of treatments to help people who find this process difficult. If you are a University of Southampton psychology student you will also receive four research credits as part of the research participation scheme. If you complete the study you can also opt into a prize draw to win one of six vouchers (1x £50, 2x £35, 3x £25), for either Amazon or Marks and Spencers.

What if I change my mind?

If at any time you change your mind, and no longer wish to take part, you can withdraw your consent without any negative consequences.

What happens to the information I give?

Any information that you give will be kept strictly confidential, and will not be released to, or viewed by, anyone other than researchers involved in this study, though it may be used to support further research. All data will be stored in an anonymous format, and results of this study will not include your name or any other identifying characteristics.

What will happen to the results of the research?

The results of the study will be written up for publication in a psychological journal. A brief summary of the findings can be made available upon request.

Who has reviewed the study?

This study has been reviewed by the School of Psychology Research Ethics Committee, University of Southampton.

Contact for further information

If you would like more information about any aspect of the study, please do not hesitate to contact Zoe McAndrews, on 07843985355, or email: zm1g14@soton.ac.uk at School of Psychology, Building 44A, Highfield Campus, University of Southampton, SO17 1BJ.

Finally, if you have any questions about your rights as a participant in this research, please contact the Chair of the Ethics Committee, School of Psychology, University of Southampton, Southampton, SO17 1BJ, Tel: 023 80593995 or email fhs-rso@soton.ac.uk

Thank you for taking the time to read this information.

Zoe McAndrews, Trainee Clinical Psychologist, University of Southampton

Dr Lusia Stopa, Professor of Clinical Psychology, University of Southampton

Dr Claire Hart Lecturer in Personality/Social Psychology, University of Southampton

Dr Laura Brummer, Principal Clinical Psychologist, Southern Health NHS Foundation Trust

Appendix H Attitudes and Emotions Across the Life

Span: Debriefing Statement

Background

Research suggests that as we grow older and have different experiences, the way in which we manage our emotions may change. This study aimed to explore the way in which ageing, life experiences and personal characteristics impacts on our responses to emotions and emotional situations.

Methodology

Participants were asked to complete ten questionnaires, which aim to explore how people tend to respond to emotional events that may be stressful or difficult. Descriptive information relating to characteristics, such as age and gender, were also collected in order to describe the participants who took part.

Results

The results will be written up and submitted for publication in a peer reviewed journal. This article will summarise the overall findings of the study, but will not identify any individuals. A brief summary of the findings can be made available to participants upon request.

Your responses

If anything that you have been asked to complete as part of this study has caused you to experience any negative emotions with which you are having difficulty coping, there are a number of sources of support available to you:

1. You can contact your general practitioner (GP).
2. If you are registered with a Southampton GP, you can contact the Steps2Wellbeing IAPT service, at:

Third Floor, Grenville House, Nelson Gate, Southampton, SO15 1GX

Tel. 0800 612 7000 or 02380 272000

Email: dhc.sstw@nhs.net

Web: www.steps2wellbeing.co.uk

If you do not live in Southampton, you can find information about your local Improving Access to Psychological Therapies (IAPT) service here:

[http://www.nhs.uk/Service-Search/Psychological%20therapies%20\(IAPT\)/LocationSearch/10008](http://www.nhs.uk/Service-Search/Psychological%20therapies%20(IAPT)/LocationSearch/10008)

3. If you are a University of Southampton student, you can contact the Counselling Service at:

Enabling Services, Student Services Building 37, Highfield Campus, Southampton, SO17 1DU

Telephone: 023 8059 7726

Email: enable@soton.ac.uk

Further information

It has been suggested that the way we experience emotions changes as we get older, but is unclear whether older adults experience more or less negative emotions when compared to younger adults. A possible explanation for this change in experience is that older adults make use of different strategies to manage their emotions. Some strategies, such as suppression and avoidance, have been linked to increased experience of negative emotions, whilst others, cognitive reappraisal and acceptance for example, are said to be associated with increased wellbeing. However, it remains unclear whether this relationship is constant across the lifespan. Furthermore, it is not clear how different life experiences affect emotional experience and resilience.

The current study aimed to explore the use of such strategies in different age groups, and their relationship with psychological wellbeing. We also aimed to explore how different life experiences and personal characteristics affect this relationship.

If you are interested in reading more about emotion regulation and ageing, please see:

Brummer, L., Stopa, L. & Bucks, R. (2013). The influence of age on emotion regulation strategies and psychological distress. *Behavioural and Cognitive Psychotherapy*, 42, 668-681.

Appendix H

Gooding, P. A., Hurst, A., Johnson, J., & Tarrier, N. (2012). Psychological resilience in young and older adults. *International Journal of Geriatric Psychiatry*, 27(3), 262-270.

John, O. P., & Gross, J. J. (2004). Healthy and Unhealthy emotion regulation: Personality processes, individual differences, and lifespan development. *Journal of Personality*, 72, 1301-1334.

Finally, if you have any questions about your rights as a participant in this research, or you feel that you have been placed at risk, please contact the Chair of the Ethics Committee, School of Psychology, University of Southampton, Southampton, SO17 1BJ, Tel: 023 80593995 or email fshs-rso@soton.ac.uk.

If you know of anyone aged between 18-30, 40-55 or over 65 years of age who might be interested in taking part in this research, you can share the following link with them:

<https://www.isurvey.soton.ac.uk/23779>

Thank you for taking the time to read this information.

Zoe McAndrews, Trainee Clinical Psychologist, University of Southampton

Dr Lusia Stopa, Professor of Clinical Psychology, University of Southampton

Dr Claire Hart, Lecturer in Personality/Social Psychology, University of Southampton

Dr Laura Brummer, Principal Clinical Psychologist, Southern Health NHS Foundation
Trust

Appendix I The Southampton Acceptance Scale (SAS)

We are interested in how people regulate and manage their emotions. Below are a list of statements. Although some of these statements may seem similar to one another, they differ in important ways. For each item please rate how true this is for you, using the following scale:

	1	2	3	4	5	6	7
	Never True	Very Rarely True	Rarely True	Sometimes True	Often True	Almost Always True	Always True
1. I try to put problems out of my mind	1	2	3	4	5	6	7
2. I allow myself to experience all of my emotions	1	2	3	4	5	6	7
3. I am willing to fully experience all thoughts that come into my mind	1	2	3	4	5	6	7
4. I allow myself to experience unpleasant emotions, even if they get worse	1	2	3	4	5	6	7
5. I tell myself that I should not have certain thoughts	1	2	3	4	5	6	7
6. I try to distract myself when I feel unpleasant emotions	1	2	3	4	5	6	7
7. I allow myself to experience unpleasant thoughts, even if they get worse	1	2	3	4	5	6	7
8. I work hard to keep out unpleasant thoughts	1	2	3	4	5	6	7
9. I allow myself to experience emotions for as long as they remain	1	2	3	4	5	6	7
10. I am willing to fully experience all emotions that arise	1	2	3	4	5	6	7
11. When I experience unpleasant emotions, I try to change or get rid of them	1	2	3	4	5	6	7
12. It is important to keep emotions under control	1	2	3	4	5	6	7

Appendix I

13. I often do things to avoid unpleasant emotions	1	2	3	4	5	6	7
14. I allow my thoughts to be there for as long as they remain	1	2	3	4	5	6	7
15. I allow my thoughts to come and go freely	1	2	3	4	5	6	7
16. I work hard to keep out unpleasant emotions	1	2	3	4	5	6	7
17. I tell myself I should not feel certain things	1	2	3	4	5	6	7
18. I welcome all emotions	1	2	3	4	5	6	7

Appendix J Acceptance and Action Questionnaire- Revised (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

Appendix K Emotion Regulation Questionnaire (ERQ)

We would like to ask you some questions about your emotional life, in particular, how you control (that is, regulate and manage) your emotions.

We are interested in two aspects of your emotional life. One is your emotional experience, or what you feel like inside. The other is your emotional expression, or how you show your emotions in the way you talk, gesture, or behave.

Although some of the following questions may seem similar to one another, they differ in important ways.

For each item, please answer using the following scale:

1-----2-----3-----4-----5-----6-----7

**Strongly
disagree**

Neutral

**Strongly
agree**

1.	When I want to feel more <i>positive</i> emotion (such as joy or amusement), I <i>change what I'm thinking about</i> .	1	2	3	4	5	6	7
2.	I keep my emotions to myself.	1	2	3	4	5	6	7
3.	When I want to feel less <i>negative</i> emotion (such as sadness or anger), I <i>change what I'm thinking about</i> .	1	2	3	4	5	6	7
4.	When I am feeling <i>positive</i> emotions, I am careful not to express them.	1	2	3	4	5	6	7
5.	When I'm faced with a stressful situation, I make myself <i>think about it</i> in a way that helps me stay calm.	1	2	3	4	5	6	7
6.	I control my emotions by <i>not expressing them</i> .	1	2	3	4	5	6	7
7.	When I want to feel more <i>positive</i> emotion, I <i>change the way I'm thinking</i> about the situation.	1	2	3	4	5	6	7
8.	I control my emotions by <i>changing the way I think</i> about the situation I'm in.	1	2	3	4	5	6	7
9.	When I am feeling <i>negative</i> emotions, I make sure not to express them.	1	2	3	4	5	6	7
10.	When I want to feel less <i>negative</i> emotion, I <i>change the way I'm thinking</i> about the situation.	1	2	3	4	5	6	7

Appendix L Positive and Negative Affect Scale

(PANAS)

This scale consists of a number of words that describe different feelings and emotions.

Please read each item and then select the appropriate answer to indicate to what extent you have felt this way in the past two weeks.

Use the following scale to record your answers:

1 - very slightly or not at all 2 - a little 3 - moderately 4 - quite a bit 5 – extremely

___ Interested

___ Irritable

___ Distressed

___ Alert

___ Excited

___ Ashamed

___ Upset

___ Inspired

___ Strong

___ Nervous

___ Guilty

___ Determined

___ Scared

___ Attentive

___ Hostile

___ Jittery

___ Enthusiastic

___ Active

___ Proud

___ Afraid

Appendix M Short Warwick Edinburgh Mental Well-being Scale (SWEMWBS)

Below are some statements about feelings and thoughts.

Please tick the box that best describes your experience of
each over the last 2 weeks

STATEMENTS	None of the time	Rarely	Some of the time	Often	All of the time
I've been feeling optimistic about the future	1	2	3	4	5
I've been feeling useful	1	2	3	4	5
I've been feeling relaxed	1	2	3	4	5
I've been dealing with problems well	1	2	3	4	5
I've been thinking clearly	1	2	3	4	5
I've been feeling close to other people	1	2	3	4	5
I've been able to make up my own mind about things	1	2	3	4	5

Appendix N SWEMWBS Conversion Table

Raw score to metric score conversion table for SWEMWBS.

Raw Score	Metric Score
7	7.00
8	9.51
9	11.25
10	12.40
11	13.33
12	14.08
13	14.75
14	15.32
15	15.84
16	16.36
17	16.88
18	17.43
19	17.98
20	18.59
21	19.25
22	19.98

Appendix N

23	20.73
24	21.54
25	22.35
26	23.21
27	24.11
28	25.03
29	26.02
30	27.03
31	28.13
32	29.31
33	30.70
34	32.55
35	35.00

Stewart-Brown *et al.* *Health and Quality of Life Outcomes* 2009 7:15 doi:10.1186/1477-7525-7-15

Appendix O **Balanced Inventory of Desired Responding – Short Form (BIDR-16), Impression Management Subscale**

Please read the statements below, and indicate how much you agree or disagree with each one:

1 = *totally disagree* 2 3 4 5 6 7 8 = *totally agree*

___ I sometimes tell lies if I have to.

___ I never cover up my mistakes.

___ There have been occasions when I have taken advantage of someone.

___ I sometime try to get even rather than forgive and forget.

___ I have never said something bad about a friend behind his or her back.

___ When I hear people talking privately, I avoid listening.

___ I never take things that don't belong to me.

___ I don't gossip about other people's business.

List of References

- Acceptance. [Def. 3]. (2010). *Oxford Dictionary of English*. (A. Stevenson, Ed.). Oxford University Press. <https://doi.org/10.1093/acref/9780199571123.001.0001>
- Allen, V. C., & Windsor, T. D. (2017). Age differences in the use of emotion regulation strategies derived from the process model of emotion regulation: A systematic review. *Ageing & Mental Health*. <https://doi.org/10.1080/13607863.2017.1396575>
- Andrei, F., Vesely, A., & Siegling, A. B. (2016). An examination of concurrent and incremental validity of four mindfulness scales. *Journal of Psychopathology and Behavioral Assessment*, 38(4), 559–571. <https://doi.org/10.1007/s10862-016-9546-x>
- Arbuckle, J. L. (2017). *Amos* (Version 25.0) [Computer Program]. Chicago: IBM SPSS.
- Baer, R. A., Carmody, J., & Hunsinger, M. (2012). Weekly change in mindfulness and perceived stress in a Mindfulness-Based Stress Reduction program. *Journal of Clinical Psychology*, 68(7), 755–765. <https://doi.org/10.1002/jclp.21865>
- Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky inventory of mindfulness skills. *Assessment*, 11(3), 191–206. <https://doi.org/10.1177/1073191104268029>
- Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, 13(1), 27–45. <https://doi.org/10.1177/1073191105283504>
- Baer, R. A., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., ... Williams, J. M. G. (2008). Construct validity of the Five Facet Mindfulness Questionnaire in

List of References

meditating and nonmeditating samples. *Assessment*, 15(3), 329–342.

<https://doi.org/10.1177/1073191107313003>

Barker, C., Pistrang, N., & Elliott, R. (2002). *Research methods in clinical psychology: An introduction for students and practitioners*. (2nd ed.). West Sussex, UK: John Wiley & Sons Ltd.

Benjamini, Y., & Yekutieli, D. (2001). The control of the False Discovery Rate in multiple testing under dependency. *The Annals of Statistics*, 29(4), 1165–1188.

<https://doi.org/10.1214/aos/1013699998>

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588–606.

<https://doi.org/10.1037/0033-2909.88.3.588>

Blackledge, J. T., & Hayes, S. C. (2001). Emotion regulation in acceptance and commitment therapy. *Journal of Clinical Psychology*, 57(2), 243–255.

[https://doi.org/10.1002/1097-4679\(200102\)57:2<243::AID-JCLP9>3.0.CO;2-](https://doi.org/10.1002/1097-4679(200102)57:2<243::AID-JCLP9>3.0.CO;2-X)

X

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.

<https://doi.org/10.1016/j.beth.2011.03.007>

Byrne, B. M. (2010). *Structural Equation Modeling with AMOS: Basic concepts, applications and programming*. New York, New York: Routledge.

- 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.
<https://doi.org/10.1177/1073191107311467>
- Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology*, *56*(2), 267–283. <https://doi.org/10.1037/0022-3514.56.2.267>
- Christopher, M. S., Neuser, N. J., Michael, P. G., & Baitmangalkar, A. (2012). Exploring the psychometric properties of the five facet mindfulness questionnaire. *Mindfulness*, *3*(2), 124-131. DOI. 10.1007/s12671-011-0086-x
- Cloitre, M., Stolbach, B. C., Herman, J. L., Van Der Kolk, B., Pynoos, R., Wang, J., & Petkova, E. (2009). A developmental approach to complex PTSD: Childhood and adult cumulative trauma as predictors of symptom complexity. *Journal of Traumatic Stress*, *22*(5), 399–408. <https://doi.org/10.1002/jts.20444>
- Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British journal of clinical psychology*, *43*(3), 245-265.
- Curtiss, J., & Klemanski, D. H. (2014). Factor analysis of the Five Facet Mindfulness Questionnaire in a heterogeneous clinical sample. *Journal of Psychopathology and Behavioral Assessment*, *36*(4), 683–694. <https://doi.org/10.1007/s10862-014-9429-y>
- DeVellis, R. F. (2003). *Scale development: Theory and applications (Second Edition)*. Thousand Oaks, CA, US: Sage Publications.
- de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in Medicine - A practical guide*. New York: Cambridge University Press.

List of References

- 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. <https://doi.org/10.1016/j.jclinepi.2013.12.002>
- 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. https://doi.org/10.1207/s15328007sem1303_6
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics (Fourth Edition)*. London: Sage.
- Francis, A. W., Dawson, D. L., & Golijani-Moghaddam, N. (2016). The development and validation of the Comprehensive assessment of Acceptance and Commitment Therapy processes (CompACT). *Journal of Contextual Behavioral Science*, *5*(3), 134–145. <https://doi.org/10.1016/j.jcbs.2016.05.003>
- 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. <https://doi.org/10.1037/a0023242>
- 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. <https://doi.org/10.1016/j.beth.2013.09.001>
- Goldberg, S. B., Wielgosz, J., Dahl, C., Schuyler, B., MacCoon, D. S., Rosenkranz, M., ... Davidson, R. J. (2016). Does the Five Facet Mindfulness Questionnaire measure what we think it does? Construct validity evidence from an active controlled

- randomized clinical trial. *Psychological Assessment*, 28(8), 1009–1014.
<https://doi.org/10.1037/pas0000233>
- Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, 26(1), 41–54. <https://doi.org/10.1007/s10862-008-9102-4>
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, 31(4), 337–350. <https://doi.org/10.1007/s10654-016-0149-3>
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85(2), 348–362. <https://doi.org/10.1037/0022-3514.85.2.348>
- Gu, J., Strauss, C., Crane, C., Barnhofer, T., Karl, A., Cavanagh, K., & Kuyken, W. (2016). Examining the factor structure of the 39-item and 15-item versions of the Five Facet Mindfulness Questionnaire before and after mindfulness-based cognitive therapy for people with recurrent depression. *Psychological Assessment*, 28(7), 791–802. <https://doi.org/10.1037/pas0000263>
- Harris, R. (2009). *ACT made simple*. Oakland, CA: New Harbinger Publications, Inc.
- Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The Balanced Inventory of Desirable Responding Short Form (BIDR-16). *SAGE Open*, 5(4), 1–9. <https://doi.org/10.1177/2158244015621113>

List of References

- Hawley, L. L., Rogojanski, J., Vorstenbosch, V., Quilty, L. C., Laposa, J. M., & Rector, N. A. (2017). The structure, correlates, and treatment related changes of mindfulness facets across the anxiety disorders and obsessive compulsive disorder. *Journal of Anxiety Disorders, 49*, 65–75. <https://doi.org/10.1016/j.janxdis.2017.03.003>
- Hayes, S. C. (2004). Acceptance and Commitment Therapy, Relational Frame Theory, and the third wave of Behavioral and Cognitive Therapies. *Behavior Therapy, 35*, 639–665. [https://doi.org/10.1016/S0005-7894\(04\)80013-3](https://doi.org/10.1016/S0005-7894(04)80013-3)
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition*. New York: Plenum Press.
- Hayes, S. C., Levin, M. E., Plumb-Villardaga, 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. <https://doi.org/10.1016/J.BETH.2009.08.002>
- 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. <https://doi.org/10.1016/j.brat.2005.06.006>
- 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., Strosahl, K., Wilson, K. G., Bissett, R. T., Pistorello, J., Toarmino, D., ...
McCurry, S. M. (2004). Measuring experiential avoidance: A preliminary test of a working model. *The Psychological Record, 54*(4), 553–578.
<https://doi.org/10.1007/BF03395492>

- Hayes, S. C., Wilson, K. W., 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. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8991302>
- Higgins, J. P. T., & Green, S. (editors). (2011). *Cochrane Handbook for Systematic
Reviews of Interventions*, Version 5.1.0 [updated March 2011]. The Cochrane
Collaboration. Available from www.handbook.cochrane.org.
- Hofmann, S. G., & Kashdan, T. B. (2010). The Affective Style Questionnaire:
Development and psychometric properties. *Journal of Psychopathology and
Behavioral Assessment*, *32*(2), 255–263. <https://doi.org/10.1007/s10862-009-9142-4>
- Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long
(Eds.), *Testing Structural Equation Models* (pp. 294–396). London: Sage
Publications.
- Kline, P. (2000). *The Handbook of psychological testing (Second Edition)*. *Personality and
Individual Differences*. London: Routledge. [https://doi.org/10.1016/S0191-
8869\(96\)90047-1](https://doi.org/10.1016/S0191-8869(96)90047-1)
- Koole, S. (2009). The psychology of emotion regulation: An integrative review. *Cognition
and Emotion*. <https://doi.org/10.1080/02699930802619031>
- Kortte, K. B., Veiel, L., Batten, S. V., & Wegener, S. T. (2009). Measuring avoidance in
medical rehabilitation. *Rehabilitation Psychology*, *54*(1), 91–98.
<https://doi.org/10.1037/a0014703>
- Lavender, J. M., Tull, M. T., DiLillo, D., Messman-Moore, T., & Gratz, K. L. (2017).
Development and validation of a state-based measure of emotion dysregulation.
Assessment, *24*(2), 197–209. <https://doi.org/10.1177/10731911155601218>

List of References

- Lazarus, R. S., & Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York, NY: Springer.
- Leahy, L. L., Tirsch, D., & Napolitano, L. A. (2011). *Emotion regulation in psychotherapy: A practitioner's guide*. New York, New York: The Guildford Press.
- Linehan, M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. New York, NY: The Guilford Press.
- Little, R. J. A., & Rubin, D. B. (2014). *Statistical analysis with missing data (Second Edition)*. Hoboken, NJ: John Wiley & Sons, Inc.
- Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285–300. <https://doi.org/10.1037/a0033266>
- 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. <https://doi.org/10.1016/j.jpain.2013.02.009>
- McCracken, L. M., Vowles, K. E., & Eccleston, C. (2004). Acceptance of chronic pain: Component analysis and a revised assessment method. *Pain*, *107*(1), 159–166. <https://doi.org/10.1016/j.pain.2003.10.012>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine*, *6*, (7), e1000097. doi:10.1371/journal.pmed1000097
- Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D., ...de Vet, H. (2010a). International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. *Journal of Clinical Epidemiology*, *63*, 737-745.

- Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D., ... de Vet, H. (2010b). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research, 19*, 539-549.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA, US: Sage Publications.
- Niebuhr, R., & Brown, R. M. (1986). *The essential Reinhold Niebuhr : Selected essays and addresses*. New Haven: Yale University Press.
- Palladino, C. L., Ange, B., Richardson, D. S., Casillas, R., Decker, M., Gillies, R. A., ... Stepleman, L. (2013). Measuring psychological flexibility in medical students and residents: A psychometric analysis. *Medical Education Online, 18*(1), 20932. <https://doi.org/10.3402/>
- Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ, 316*(7139), 1236–1238. <https://doi.org/10.1136/bmj.316.7139.1236>
- 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.
- Pike, N. (2011). Using False Discovery Rates for multiple comparisons in ecology and evolution. *Methods in Ecology and Evolution, 2*(3), 278–282. <https://doi.org/10.1111/j.2041-210X.2010.00061.x>
- Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis. *Personality and Social Psychology Bulletin, 28*(12), 1629–1646. <https://doi.org/10.1177/014616702237645>

List of References

- Sandoz, E. K., Wilson, K. G., Merwin, R. M., & Kate Kellum, K. (2013). Assessment of body image flexibility: The Body Image-Acceptance and Action Questionnaire. *Journal of Contextual Behavioral Science*, 2(1–2), 39–48.
<https://doi.org/10.1016/j.jcbs.2013.03.002>
- Scott, W., McCracken, L. M., & Norton, S. (2016). A Confirmatory Factor Analysis of Facets of Psychological Flexibility in a Sample of People Seeking Treatment for Chronic Pain. *Annals of Behavioral Medicine*, 50(2), 285–296.
<https://doi.org/10.1007/s12160-015-9752-x>
- Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. New York, NY: Guilford Press.
- Shawyer, F., Ratcliff, K., Mackinnon, A., Farhall, J., Hayes, S. C., & Copolov, D. (2007). The Voices Acceptance and Action Scale (VAAS): Pilot data. *Journal of Clinical Psychology*, 63(6), 593–606. <https://doi.org/10.1002/jclp.20366>
- Siegling, A. B., & Petrides, K. V. (2016). Zeroing in on mindfulness facets: Similarities, validity, and dimensionality across three independent measures. *PLoS ONE*, 11(4), e0153073. <https://doi.org/10.1371/journal.pone.0153073>
- Singer, E., & Couper, M. P. (2008). Do incentives exert undue influence on survey participation? Experimental evidence. *Journal of Empirical Research on Human Research Ethics*, 3(3), 49–56. <https://doi.org/10.1525/jer.2008.3.3.49>
- Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): A Rasch analysis using data from the Scottish Health

- Education Population Survey. *Health and Quality of Life Outcomes*, 7(1), 15.
<https://doi.org/10.1186/1477-7525-7-15>
- Streiner, G. R., Norman, D. L., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use (Fifth Edition)* (Fifth). Oxford University Press.
- Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics: Pearson New International Edition (Sixth Edition)*. Harlow, Essex: Pearson Education Ltd.
- Teasdale, J. D. (1999). Metacognition, mindfulness and the modification of mood disorders. *Clinical Psychology and Psychotherapy*, 6(2), 146–155.
[https://doi.org/10.1002/\(SICI\)1099-0879\(199905\)6:2<146::AID-CPP195>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1099-0879(199905)6:2<146::AID-CPP195>3.0.CO;2-E)
- Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., ... de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>
- Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 18(8), 1115–23.
<https://doi.org/10.1007/s11136-009-9528-5>
- Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W. J. G., Bouter, L. M., & De Vet, H. C. W. (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. <https://doi.org/10.1007/s11136-011-9960-1>

List of References

- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.
<https://doi.org/10.1037/0022-3514.54.6.1063>
- Watson-Singleton, N. N., Walker, J. H., LoParo, D., Mack, S. A., & Kaslow, N. J. (2018). Psychometric evaluation of the Five Facet Mindfulness Questionnaire in a clinical sample of African Americans. *Mindfulness*, *9*(1), 312–324.
<https://doi.org/10.1007/s12671-017-0776-0>
- Wegner, D. M., & Zanakos, S. (1994). Chronic Thought Suppression. *Journal of Personality*, *62*(4), 615–640. <https://doi.org/10.1111/j.1467-6494.1994.tb00311.x>
- Wells, A. (2009). *Metacognitive therapy for anxiety and depression*. New York, NY: Guilford Press.
- Westbrook, D. (2014). The Central Pillars of CBT. In A. Whittington & N. Grey (Eds.), *How to become a more effective CBT Therapist: Mastering metacompetencies in clinical practice* (pp. 17–30). Chichester, West Sussex: John Wiley & Sons Ltd.
Retrieved from <https://ebookcentral.proquest.com/lib/soton-ebooks/reader.action?docID=1676371&ppg=21>
- Westin, V., Hayes, S. C., & Andersson, G. (2008). Is it the sound or your relationship to it? The role of acceptance in predicting tinnitus impact. *Behaviour Research and Therapy*, *46*(12), 1259–1265. <https://doi.org/10.1016/j.brat.2008.08.008>
- Williams, M. J., Dalgleish, T., Karl, A., & Kuyken, W. (2014). Examining the factor structures of the five facet mindfulness questionnaire and the self-compassion scale. *Psychological Assessment*, *26*(2), 407–418. <https://doi.org/10.1037/a0035566>

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

<https://doi.org/10.1016/j.beth.2014.07.002>

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806–838. <https://doi.org/10.1177/0011000006288127>