



University of Roehampton

PSYCHD

The Authenticity Scale as an outcome measure for psychological therapies

Mørken, Daniel

Award date:
2019

Awarding institution:
University of Roehampton

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**The Authenticity Scale as an outcome measure for
psychological therapies**

by

Daniel Martin Mørken, BSc

*A thesis submitted in partial fulfilment of the requirements for the degree of
PsychD in Counselling Psychology*

**Department of Psychology
University of Roehampton**

2019

Abstract

Background. Counselling and psychotherapy services are under pressure to evidence clinical effectiveness. However, many of the outcome measures in use are based on the ‘disease model’: a paradigm that is at odds with several key psychotherapy approaches, including the humanistic approach. This has led to calls for psychometrically sound outcome measures that emphasise growth and well-being. One such measure is the Authenticity Scale (AS), based on the Rogerian concept of congruence. Authenticity is a core construct within humanistic psychology and the person-centred model of psychotherapy, and the measure has been developed specifically for use as an outcome tool for counselling and psychotherapy. Despite this, no research to date has assessed the validity of the AS in this way.

Aims. The aim of this thesis is to investigate the psychometric properties of the AS, and how it functions as an outcome measure, specifically looking at its reliability, validity, acceptability, and sensitivity to change in a clinical sample.

Methods. The thesis consists of two studies. Study 1 is a systematic review of the literature regarding the AS in general populations, collating psychometric data from 60 publications into a narrative synthesis and meta-analysis. In Study 2, a total of 67 clients receiving psychotherapy completed the AS and the CORE-10 periodically over the course of their therapy. Sensitivity to change was assessed using pre- and post-therapy comparisons, and multilevel modelling.

Results. In Study 1, the AS showed excellent psychometric properties in the majority of samples, confirming its potential as a psychotherapy outcome measure. In Study 2, the AS was internally reliable in the clinical sample, and showed convergent validity with the CORE-10 measure of psychological distress. Multilevel modelling indicated that the AS was sensitive to change, with participants’ scores increasing significantly over the course of therapy. These

findings indicate that the AS may be a psychometrically valid and reliable outcome measure for use in the psychological therapies.

Table of Contents

Acknowledgements	8
Structure of the thesis	10
Chapter 1. Background.....	11
1.1. Introduction	11
1.2. Personal rationale	11
1.3. Background	13
1.3.1. The move to outcome monitoring	13
1.3.2. Why monitor outcomes?.....	15
1.3.3. Challenges of outcome monitoring	16
1.3.4. Popular outcome measures	17
1.3.4.1. The PHQ-9 and the GAD-7	18
1.3.4.2. CORE-OM/CORE-10.....	19
1.3.5. Positive psychology and humanistic psychology	20
1.3.6. Well-being measures	22
1.3.6.1. The WEMWBS	22
1.3.6.2. The Strathclyde Inventory (SI)	23
1.3.7. Authenticity	24
1.3.7.1. State and trait authenticity	25
1.3.8. Authenticity in person-centred psychology	27
1.3.9. Measuring authenticity	29
1.3.9.1. The Experienced Authenticity Measure (EAM).....	30
1.3.9.2. The Authenticity in Relationships Scale (AIRS).....	31
1.3.9.3. The Perception of False Self Scale (POFS).....	31
1.3.9.4. The Authenticity Inventory (AI)	32
1.3.10. The Authenticity Scale	33
1.3.10.1. The Authenticity Scale as an outcome measure	35
1.3.10.2. The psychometric properties of the Authenticity Scale.....	36
1.4. Conclusion	37
1.5. A priori aims	38
STUDY 1.....	39
Chapter 2. Systematic review of the psychometric properties of the Authenticity Scale, with a narrative synthesis and meta-analysis	39
2.1. Introduction	39
2.2. Methods	40

2.2.1. Aims of the systematic review and research questions	40
2.2.2. Methods of the systematic review	41
2.2.2.1. Registration of the systematic review.....	42
2.2.2.2. Eligibility criteria.....	42
2.2.2.3. Data sources and search strategy	42
2.2.2.4. Study selection.....	43
2.2.3. Methods of the narrative synthesis	44
2.2.3.1. Construct validity	44
2.2.4. Methods of the meta-analysis	45
2.2.4.1. Data analysis.....	46
2.2.4.2. Random-effects model.....	46
2.2.4.3. Calculating the combined effect under a random-effects model.....	47
2.2.4.4. Heterogeneity	48
2.2.4.5. Reliability	48
2.2.4.6. Construct validity	49
2.2.4.7. Publication bias	50
2.2.4.8. Percentage of maximum possible (POMP) scores	51
2.2.4.9. Dealing with erroneous data	52
2.3. Results.....	52
2.3.1. Description of the studies and samples	52
2.3.2. Results of the narrative synthesis	55
2.3.2.1. Concurrent validity	55
2.3.2.2. The AS as an outcome measure and with clinical samples	56
2.3.3. Results of the meta-analysis	57
2.3.3.1. Internal reliability (coefficient alpha).....	57
2.3.3.2. Scale intercorrelations	59
2.3.3.3. Mean scores	62
2.4. Discussion	65
2.4.1. Narrative synthesis discussion.....	65
2.4.1.1. Concurrent validity	65
2.4.1.2. The AS and measures of well-being.....	65
2.4.1.3. The AS predicts scores on the SWLS.....	66
2.4.1.4. The AS and measures of emotional distress.....	68
2.4.1.5. The AS and other measures of authenticity.....	69
2.4.1.6. The AS and measures of personality	71
2.4.1.7. The AS with clinical samples	73
2.4.2. Meta-analysis discussion	73
2.4.2.1. Reliability generalisation.....	74
2.4.2.2. Scale intercorrelations	76
2.4.2.3. Mean scores and distributions	77
2.4.2.4. Limitations to the meta-analysis.....	80
2.5. Conclusion and implications for Study 2.....	81
STUDY 2.....	83
Chapter 3. Methods	83
3.1. Introduction	83

3.2. Epistemology	83
3.2.1. Empiricism and positivism	84
3.2.2. Realism	85
3.2.3. Constructivism.....	86
3.2.4. Pragmatism	87
3.2.5. What is a psychological construct?	88
3.2.6. Pragmatism-realism	89
3.2.7. Conclusion	90
3.3. Aims, research questions and expected findings	90
3.3.1. Research questions	90
3.3.2. Expected findings	91
3.4. Design.....	92
3.5. Procedure	92
3.5.1. Ethical approval.....	92
3.5.2. Sample size	93
3.5.3. Recruitment	93
3.6. Participants	95
3.6.1. Counselling services	95
3.6.2. Clients	96
3.6.3. Therapists	97
3.7. Measures.....	99
3.7.1. The Authenticity Scale (AS)	99
3.7.2. Clinical Outcomes in Routine Evaluation-10 (CORE-10)	99
3.7.3. Therapy Questionnaire Helpfulness Survey (TQHS).....	100
3.7.4. Measures completed at each service.....	101
3.8. Analyses	102
3.8.1. Estimation versus null-hypothesis testing	103
3.8.2. Study-wise multiple-comparisons issue	103
3.8.3. Sample sizes	104
3.8.4. Calculating baseline, post-therapy and change scores.....	104
3.8.5. Effects of client and therapist variables on AS scores.....	105
3.8.6. Acceptability.....	106
3.8.7. Reliability	106
3.8.8. Validity	107
3.8.8.1. Concurrent validity	107
3.8.8.2. Construct validity	108
3.8.8.3. Factorial validity.....	108
3.8.9. Response distributions.....	110
3.8.9.1. Gaussian distribution	110
3.8.9.2. Floor and ceiling effects	110
3.8.10. Sensitivity to change.....	111
3.8.10.1. Estimating sensitivity	111
3.8.10.2. Reliable and clinically significant change.....	112
3.8.10.3. Comparing the sensitivity of the AS and CORE-10.....	114
3.8.10.4. Multilevel modelling	114

Chapter 5. Results	117
5.1. Acceptability	117
5.1.1. Completion rates.....	117
5.1.2. Therapy Questionnaire Helpfulness Survey (TQHS) scores.....	117
5.2. Reliability	117
5.2.1. Internal reliability (coefficient alpha).....	117
5.3. Validity	118
5.3.1. Concurrent validity.....	118
5.3.2. Construct validity (subscale intercorrelations).....	120
5.3.3. Factorial validity.....	121
5.4. Response distributions	122
5.4.1. Gaussian distribution.....	122
5.4.2. Floor/ceiling effects.....	124
5.5. Sensitivity to change	126
5.5.1. Baseline and post-therapy comparison.....	126
5.5.2. Effect size of change.....	126
5.5.3. POMP change scores.....	127
5.5.4. Clinically significant change (CSC) and reliable change (RC).....	128
5.5.5. Multi-level modelling.....	129
Chapter 6. Discussion	130
6.1. The psychometric properties of the AS	130
6.1.1. Internal reliability.....	130
6.1.2. Validity.....	131
6.1.2.1. Concurrent validity.....	131
6.1.2.2. Factorial validity.....	135
6.1.3. Summary of the psychometric properties of the AS from Studies 1 and 2.....	135
6.2. The AS as an outcome measure.....	136
6.2.1. Acceptability.....	136
6.2.2. Response distribution.....	137
6.2.3. Sensitivity to change.....	138
6.2.3.1. Clinically significant (CSC) and reliable change (RC).....	139
6.2.3.2. Comparing the sensitivity of the AS with that of the CORE-10.....	140
6.2.3.3. Change scores and construct validity.....	141
6.2.4. Summary of how the AS functions as an outcome measure.....	142
6.3. Implications for counselling psychology.....	143
6.4. Limitations	144
6.4.1. Sample size.....	144
6.5. Directions for future research	145
6.5.1. The present research as a pilot study.....	145
6.5.2. Exploring how psychotherapy impacts authenticity.....	146
6.5.3. Exploring the relationship between authenticity and distress.....	146
6.5.4. The AS, CORE-10, and relapse.....	147

Chapter 7. Conclusions	148
7.1. Summary	148
7.2. Reflexivity.....	149
References	151
Appendices	184
Appendix A: The Authenticity Scale	184
Appendix B: Permission from APA to use the Authenticity Scale	185
Appendix C: CORE-10.....	186
Appendix D: Ethics application approval.....	187
Appendix E: Recruitment email	188
Appendix F: Counselling service information sheet	189
Appendix G: Counselling service consent form.....	191
Appendix H: Therapist information sheet	192
Appendix I: Therapist consent form.....	194
Appendix J: Therapist demographics form	195
Appendix K: Client information sheet	197
Appendix L: Client consent form	199
Appendix M: Client demographics form.....	200
Appendix N: Forest and funnel plots of the coefficient alphas for the AS subscales	201
Appendix O: Forest and funnel plots of the mean scores for the AS subscales	204
Appendix P: Study 2 - Effects of client demographics	207
Gender	207
Age	208
Ethnicity	209
Education.....	209
Appendix Q: Study 2 - Effects of therapist demographics.....	210
Gender	210
Age	211
Ethnicity	211
Qualification.....	211
Professional identity	211
Therapeutic primary orientation	212

Acknowledgements

I would first like to thank my brilliant Director of Studies, Professor Mick Cooper, who has guided and helped me with this project from it was just a vague idea in my mind, to its completion. It was Mick who wrote the chapter that first gave me the idea, and it was our ensuing conversation that convinced me to pursue it. He has been a stable and consistent presence throughout this process, and always made time to discuss my questions and thoughts. The numerous times I have been overwhelmed and doubted my ability to complete this, Mick's steady reassurance, focus, and warmth has been invaluable. I feel incredibly grateful to have had him as my guide and teacher throughout this process.

I would also like to thank Professor Chris Evans, my co-supervisor and humble genius in all things statistics- and research methods-related. I have looked forward to every one of our meetings, and in the many moments of my feeling out of my depth, his support, humour, and expertise have steadied me. Though he joined my supervisory team half-way through the project, I cannot imagine doing it without him.

I would like to thank all the participating services, therapists, and clients who have given of their time, resources and energy, without which this project would not have happened.

Thank you to my mum and dad, who have shaped me more than they know. Without the love, support and pride in me that they have expressed throughout my life, I would not be me.

Finally, I would like to thank my best-friend, co-adventurer, and wife, Zoë, to whom I also dedicate this thesis. Her unwavering love and support for me has motivated me more times than she knows, and provided the perspective I have needed to complete this project. She has sacrificed more than anyone to make this dream of mine come true, and for that I will be forever grateful. I love you.

Structure of the thesis

In this counselling psychology thesis I evaluate how a psychometric measure of authenticity, the Authenticity Scale (AS; Wood, Linley, Maltby, Baliousis, & Joseph, 2008) based in humanistic and positive psychology, functions as an outcome measure for psychotherapy. It is split into two studies, the first a systematic review of the literature pertaining to the AS, and the second, an exploration of the AS as an outcome measure, analysing AS data from clients receiving psychotherapy.

In the first chapter, I provide a rationale for using the AS as an outcome measure, laying the theoretical foundation for the empirical work in the thesis, before listing my overall aims. Chapter 2 covers Study 1, a systematic review of the AS literature, including a narrative synthesis of its concurrent validity, and a meta-analysis of its psychometric properties. I conclude that the evidence suggests the AS is likely to be sufficiently psychometrically sound to be used as an outcome measure, and note that, despite this, no publications to date have used it in such a capacity. This gap in the literature is one Study 2 starts to fill.

Chapters 4 and 5 cover Study 2, exploring how the AS functions as an outcome measure in a help-seeking population. In chapter 4 I cover the methods of Study 2. I begin by delineating my epistemological position, providing a fundamental backdrop and context for the study. I go on to present the aims of the study, followed by its design and procedure, and finally introducing the analyses I will use to address my aims. Chapter 5 is the results chapter for Study 2, where I briefly, and in statistical terms, list all my findings.

In chapter 6, the discussion chapter, I bring the two studies together, discussing how the findings of the thesis as a whole have addressed my aims, and its implications for counselling psychology and future research. In chapter 7, I briefly summarise the thesis, and end by writing reflexively about my journey conducting this research.

Chapter 1. Background

The curious paradox is that when I accept myself just as I am, then I can change.

— Carl Rogers (1961), *On Becoming a Person*, p. 17

1.1. Introduction

In this first chapter, I aim to provide a background and rationale for the thesis, delineating how it fits with, and builds upon, the wider body of literature and previous research. I first give a personal rationale for why I choose to do this particular research. I then provide a background for it, beginning by reviewing the broad field of outcome measurement in counselling psychology and its history, before moving on to the relatively recent positive psychology movement and its emphasis on using well-being focused outcome measures rather than ones that measure levels of distress. I continue by introducing some well-being measures that have sprung out of this positive psychology movement, before focusing on authenticity as a construct, and ways it has been conceptualised and measured. Finally, I introduce the Authenticity Scale (Wood *et al.*, 2008), the topic of this thesis, reviewing its conceptualisation and background, and its potential use as an outcome measure in counselling psychology. I end this first chapter by introducing my two overall aims for the thesis, which are addressed in the subsequent chapters.

1.2. Personal rationale

The rationale behind my wishing to conduct this research and write this thesis are rooted in my own journey of, as Carl Rogers (1961a) might say, becoming a person. In my teenage years, I was lucky enough to have a group of friends around me with whom I felt safe enough to express what I was experiencing, which pivotally included the parts of me that I was convinced disqualified

me from the love and affection that I sought. I remember vividly the fear of letting people see those part of me I had worked hard to cover up and hide, and the deep sense of acceptance that followed being *seen*. I experienced first hand the power of authenticity, and the freedom of simply being me, unashamedly. Though I did not know there was such a profession, I trace my desire to be a counselling psychologist back to these moments: I wished to help people experience the freedom of authenticity that I had, and subsequently found myself frequently in conversations with people saying to me: 'I've never told anyone this, but...'. On the wall in my bedroom, I hung Cicero's famous quote: 'Esse quam videri' (1923, p. 205), 'To be, rather than to appear'.

My adolescent years left me convinced that authenticity was pivotal to my happiness, and when I later started my training, I was excited to find the importance of authenticity echoed by many psychologists and psychotherapists over the last century. I was particularly captivated by Carl Rogers and the person-centred approach, with Rogers seeming to hold authenticity, or congruence, as the key to optimal functioning and well-being. When I came across the Authenticity Scale (AS) in a textbook on the person-centred approach, the idea for this thesis began to take shape. I was using a distress-focused outcome measure with my own clients, yet practicing the person-centred model, and it did not quite sit right with me. I reasoned that if I must measure outcomes by quantifying experiences, then I should at least be asking questions consistent with the model I am practicing. The AS seemed the ideal measure for this, and so I was surprised when I was unable to find any research investigating its use as an outcome measure. After one of his lectures, I decided to voice my thoughts with Professor Mick Cooper, whom I greatly admired and knew was interested in outcome research. That brief conversation further convinced me to pursue these ideas, and I was thrilled when Mick later agreed to be my supervisor for this project.

Thus, I have conducted this research utterly convinced that authenticity is one of the pivotal keys to well-being, and that the AS could be a valuable outcome measure for counselling

psychology. Though this belief has motivated me when the work has been heavy or uninspiring, I am aware that it has also left me biased, hoping to find that my research confirms my belief in the importance of authenticity, and the usefulness of the AS. I have had to bear in mind that I view my research through these lenses, and my supervisors have been invaluable in helping me recognise when my biases have impacted my research decisions. On reflection, I believe I have completed this research with the conviction that whether it confirms my original beliefs or not, this is not just a thesis to get me across the finish line of qualification, but that it can greatly benefit the field of counselling psychology.

1.3. Background

1.3.1. The move to outcome monitoring

In 1952, as psychotherapy was rapidly growing to meet demands in the aftermath of World War II, Hans Eysenck published a study claiming that though about two thirds of ‘neurotics’ substantially improved after two years of psychotherapy, a similar proportion not receiving therapy improved at the same rate. Up to this point, the emphasis in psychotherapy research had largely been on the components and process of therapy, but the Eysenck (1952) study led to a new focus on outcome research, seeking to establish firmly the effectiveness of psychotherapy. This new emphasis provided a major impetus for the use of outcome measures, providing empirical and quantitative data by which psychological therapies could be evaluated (Hill & Corbett, 1993). As a result, there has been a steadily growing body of evidence attesting to the effectiveness of psychotherapy as a whole, as well as individual therapies and interventions (e.g. Cooper, 2008; Lambert, 2013, Lambert & Vermeersch, 2002; Leichenring & Rabung, 2008; Smith, Glass, & Miller, 1980).

The past two decades have seen outcome monitoring gain further momentum, as psychological services have come under increasing pressure to demonstrate clinical effectiveness as well as cost containment, both in the private and public sectors (Corrie, 2010). From this climate of evaluation and justification, *evidence-based practice* (Department of Health, 1996, 1997; Evidence-Based Medicine Working Group, 1992) has emerged as a framework guiding the delivery of psychological services, placing evidence in a key position within the field of counselling psychology (BPS, 2017; HCPC, 2015). This has further increased the demand for empirical research on the effectiveness of psychological therapies, and by extension increased the pressure to use standardised outcome measures to assess interventions (Tarescavage & Ben-Porath, 2014).

Building on this evidence-based practice framework, the concept of *routine outcome monitoring* (ROM) has developed, essentially the regular measurement of client variables, such as symptom severity or well-being, over the course of therapy (Boswell, Kraus, Miller, & Lambert, 2015; Carlier, Meuldijk, Van Vliet, Van Fenema, & Van Der Wee, 2012; Carlier & van Eeden, 2017, Wampold, 2015). ROM was first suggested by Howard, Moras, Brill, Martinovich, and Lutz (1996), and takes outcome measurement a step further: rather than just implementing a measure pre- and post therapy, it involves monitoring a client's progress over the course of therapy as well, typically completing measures sessionally. In addition to monitoring client progress, a lot of research has also looked at the effects of feeding this information back to the therapist, or to both the therapist and client, often termed ROM with feedback (Wampold, 2015). ROM has gained momentum over the past fifteen years (Boswell *et al.*, 2015), and in the UK, ROM with feedback was included as a component in the Improving Access to Psychological Therapies initiative (IAPT; Clark, Layard, Smithies, Richards, Suckling, & Wright, 2009), whilst in the US, the APA

Presidential Task Force on Evidence-Based Practice (2006) included it is a vital element of evidence-based practice.

1.3.2. Why monitor outcomes?

More than two decades ago, the American psychologist Paul Clement (1994) argued all therapists should be asked ‘Are you any good?’ (p.173). He monitored the progress of every client he had seen since he began practicing in 1966, 26 years of quantitative data, and held that all therapists have an ethical obligation to use empirical data to direct their practice and answer the question ‘with whom do I achieve the best therapeutic outcomes?’.

Twenty-five years on from Clement (1994), outcome monitoring has become a central and contentious topic in counselling psychology (see Boswell *et al.*, 2015). There is a growing body of evidence suggesting there to be important clinical benefits to outcome monitoring, particularly concerning ROM with feedback (e.g. Gondek, Edbrooke-Childs, Fink, Deighton, & Wolpert, 2016, Lambert & Shimokawa, 2011; Wampold, 2015), though there are also studies that have failed to find positive effects (e.g. de Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012; Rise, Eriksen, Grimstad, & Steinsbekk, 2016). A recent meta-analysis (Lambert, Whipple, & Kleinstäuber, 2018) reviewed 24 studies looking at the effect of feedback from outcome measures on treatment effectiveness, and found that in two-thirds of the studies, ROM-assisted psychotherapy had better outcomes than therapy-as-usual conducted by the same clinicians, with effect sizes ranging from small to moderate. ROM with feedback was found to be most beneficial for not-on-track patients.

In addition to the clinical benefits of outcome monitoring, standardised outcome measurement allows for the collection and collation of outcome data, and is thus the cornerstone of the growing body of practice-based evidence demonstrating the effectiveness of psychological

therapies (Barkham, Hardy, & Mellor-Clark, 2010). As well as enabling quantitative psychological research, furthering our understanding of the complex nature of psychotherapy and change processes (Erbes *et al.*, 2004), the present climate of high accountability (BPS, 2017; HCPC, 2015) means that evidencing treatment effectiveness is important to both services and individual practitioners, and is often key to procuring funding and support, both in the public and private sectors (Corrie, 2010; Ogles, Lambert, & Fields, 2002).

1.3.3. Challenges of outcome monitoring

Despite the benefits to employing outcome measures in clinical practice (i.e. Knaup, Koesters, Schoefer, Becker, Puschner, 2009; Lambert, *et al.*, 2018), studies have repeatedly found there is a significant research-practice gap, with a large proportion of practitioners not monitoring outcomes (Bickman *et al.*, 2000; Boswell *et al.*, 2015; Hatfield & Ogles, 2004, 2007; Ionita & Fitzpatrick, 2014; Phelps, Eisman, & Kohout, 1998; Zimmerman & McGlinchey, 2008). Fifteen years ago, Hatfield and Ogles (2004, 2007) surveyed psychologists in the US ($N = 994$) and found that 63% of respondents did not gather any outcome data. Zimmerman and McGlinchey (2008) found similar results for American psychiatrists ($N = 314$), with over 60% rarely or never implementing outcome measures, whilst in a more recent survey of Canadian psychologists ($N = 1668$), only 12.1% of respondents used outcome measures (Ionita & Fitzpatrick, 2014). Hatfield and Ogles (2007) found that practical issues, such as the added burden on clients and therapists, time constraints, and financial cost, were the most commonly cited reasons for psychologists not using outcome measures. Another issue commonly raised was around the utility of outcome measurement, with psychologists not believing it to be relevant or helpful to practice. Further to this, insight-oriented psychologists were less likely to use outcome measurement than cognitive or behavioural therapists, and the authors concluded this may be due to these practitioners believing

that the aims of treatment cannot be easily measured in the form of overt constructs such as symptoms or behaviours. Though perhaps outdated, these findings highlight some of the limitations with outcome measurement as seen by practitioners: the added administrative burden, as well as the philosophical issues around using overt constructs to quantify a clients' progress or well-being (Boswell *et al.*, 2015; Hatfield & Ogles, 2004, 2007). Addressing the latter issue, Hatfield and Ogles (2007) highlight the need for outcome measures that are more philosophically aligned to specific psychotherapeutic orientations.

The above studies were conducted in the US and Canada, and may not be applicable to the UK. However, to date there is very limited research on the use of outcome measures in the UK. The most recent study was conducted by Landy *et al.* (2013), who surveyed clinicians across three mental health trusts in London and their IAPT services ($N = 136$). Contrary to the North American studies, they found that 91% of their respondents used outcome measures in their practice. However, their sample is both relatively small and homogenous, and thus the generalisability of the survey is questionable. Interestingly, Landy *et al.* (2013) found that therapist attitudes towards ROM varied greatly depending on modality, with CBT practitioners more favourably inclined than psychodynamic psychotherapists, echoing the findings of Hatfield and Ogles (2007).

1.3.4. Popular outcome measures

Though there is no recent record of how many different outcome measures are in use, in their survey of American psychologists, Hatfield and Ogles (2004) found that there were more than 200 different standardised measures used by the 324 respondents who routinely measured outcomes. They also found that a few measures were employed by a large percentage of respondents, such as the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, &

Erbaugh, 1961) and the Global Assessment Scale (GAS; Endicott, Spitzer, Fleiss, & Cohen, 1976) (45.3% and 23.0% respectively).

There have been no recent studies surveying the use of outcome measures in the UK, but different organisations, such as IAPT (NHS Digital, 2016) and the Royal College of Psychiatrists (RCPsych; Apostolou, Ward, & Yakeley, 2016) have recommendations for which measures to use. From these recommendations I have chosen three examples that I will briefly review, before contrasting them with two measures that have emerged from positive psychology. The Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001) and Generalized Anxiety Disorder scale (GAD-7; Spitzer, Kroenke, Williams, & Löwe, 2006) are the default measures for assessing depression and anxiety in IAPT (NHS Digital, 2016), and the Clinical Outcomes in Routine Evaluation outcome measure (CORE-OM; Evans *et al.*, 2000) is recommended by the RCPsych and also used by Child and Young People's IAPT (CYP-IAPT; Wolpert, Jacob, Napoleone, Whale, Calderon, & Edbrooke-Childs, 2016).

1.3.4.1. The PHQ-9 and the GAD-7

The PHQ-9 and GAD-7 are among the most used scales that aim to measure depression and anxiety, respectively (Kroenke *et al.*, 2016). In a sample of 6000, the PHQ-9 had internal reliability (alpha) between .86 and .89 (Kroenke *et al.*, 2001), and the GAD-7 had a reported alpha of .92 in a sample of 2740 patients (Kroenke, Spitzer, Williams, & Löwe, 2010). A systematic review of the literature (Kroenke *et al.*, 2010) identified 561 publications reporting data on the PHQ (PHQ-9, PHQ-8, PHQ-2, PHQ-15) and GAD-7 scales, also reporting they had been translated into more than 80 languages, and included in many clinical practice guidelines worldwide. In the UK they have been widely employed in the NHS, and in IAPT (NHS Digital, 2016). These measures are short (9 and 7 items, respectively), unifocal and condition-specific,

purely assessing direct symptoms of anxiety and depression as defined by the DSM-IV (American Psychiatric Association, 1994). It is worth noting that both these measures are owned by the drug company Pfizer Incorporated.

The PHQ-9 and the GAD-7 both exemplify the illness-focused, medical view of psychological distress. This paradigm, often termed *the medical model* or *disease model* (Joseph & Murphy, 2013), pathologises certain human experiences, such as anxiety or depression, conceptualising them as illnesses, and psychotherapy as a medical treatment that will cure them (Wampold & Imel, 2015). Outcome measures in this paradigm aim to measure the extent to which a client is suffering, and the extent to which treatment has alleviated that suffering. Mainstream psychology has traditionally aligned itself with this model, and thus its emphasis has largely been on illness and dysfunction, and the aim of psychological intervention has been the reduction of distress, or the ‘curing’ of illness (Joseph & Murphy, 2013; Joseph & Wood, 2010). As a result, outcome measurement has largely been focused on disorder and pathology, measuring the extent to which a client is suffering, and the extent to which treatment has alleviated that suffering.

1.3.4.2. CORE-OM/CORE-10

The CORE-OM (Evans *et al.*, 2000) is another measure that is widely used by both private and public sector psychological services in the UK (Barkham *et al.*, 2013; ‘CORE Users’, n.d.). However, it diverges from the PHQ-9 and GAD-7 in that it is pantheoretical, aiming to neither align itself with the medicalised view of psychological distress, nor a positive psychology view discussed in later sections. The CORE-OM was developed to meet the need for a short and non-proprietary measure, that could be used widely by both clinicians and researchers, regardless of therapeutic modality, and that was sensitive to both high and low distress intensity (Barkham *et al.*, 1998; Evans *et al.*, 2002). Being a generic measure of psychological distress, not specific to

any condition or disorder, CORE has a much wider focus than the PHQ-9 (Kroenke *et al.*, 2001) and the GAD-7 (Spitzer *et al.*, 2006). The CORE-OM is a 34-item measure, with items split into four subscales, *subjective well-being* (4 items), *symptoms* (12 items measuring symptoms of anxiety, depression, psychological and trauma), *functioning* (12 items measuring general functioning, and close relationships and social relationships) and *risk* (6 items measuring risk to self and others). Several studies have explored its validity, and Evans *et al.* (2002) reported coefficient alphas of .94 in large clinical ($n = 713$) and non-clinical ($n = 1009$) samples. A psychometrically comparable shorter version, the CORE-10 (Barkham, *et al.*, 2013), has since been developed, pulling 10 representative items from the original 34. Barkham and colleagues (2013) found it correlated at .94 with the CORE-OM in a clinical sample and at .92 in a non-clinical, and reported internal reliability of .90.

1.3.5. Positive psychology and humanistic psychology

The past two decades have seen the emergence and growing momentum of the positive psychology perspective, placing emphasis on positive functioning, rather than psychological distress and the alleviation of suffering (Seligman, 1999; Seligman & Csikszentmihalyi, 2000). Positive psychology emerged as a response to the strong emphasis on the understanding and treating of mental illness within psychology (Waterman, 2013). In Seligman's (1999) Presidential Address to the APA he argued that since World War II, psychology has largely focused on 'repairing damage within a disease model of human functioning. Such almost exclusive attention to pathology neglects the flourishing individual and the thriving community', going on to say that 'When we became solely a healing profession, we forgot our larger mission: that of making the lives of all people better' (p. 2).

However, Joseph and Murphy (2013) hold the view that because modern positive psychology is grounded in mainstream psychology, it too has largely held to a disease model as its paradigm. Joseph and Linley (2006) explain that positive psychology follows a continuum model of human functioning, seeing positive and negative experience as opposite ends of this continuum, and thus interventions that decrease negative experience also increase positive experience, and vice versa. Thus, though positive psychology attempts to shift the emphasis to positive functioning, they argue that it still pathologises the negative end of the human experience spectrum. Joseph and Murphy (2013) argue that because of this, positive psychology ‘on the one hand implicitly condones the pathologization of human experience that on the other it was created in response to’ (p. 34).

Positive psychology builds upon humanistic psychology, and the work of Carl Rogers and Abraham Maslow (Seligman, Steen, Park & Peterson, 2005). Indeed, humanistic psychology is a positive psychology, in the sense that it too is primarily concerned with optimal functioning and potentiality (Schneider, 2011). Put briefly, humanistic psychology is based on the key construct of the *actualising tendency*, an innate, biological motivation moving us towards growth and autonomy (Rogers, 1963). When social-environmental conditions are optimal, the actualising tendency drives us towards the ideal of being *fully functioning*, according to Rogers (1963). However, when conditions are less than optimal, this tendency gets skewed and the individual self-actualises in ways consistent with what they believe their social environment expects from them. Thus, humanistic psychology understands suboptimal human functioning as thwarted human potential, rather than as illness, disorder or pathology (Joseph & Murphy, 2013). It is unsurprising, then, that both Rogers and Maslow were sharp critics of the disease model as applied to psychology, believing psychology would then serve to help people on the one hand, whilst

alienating and damaging them at the same time, by pathologising parts of their experience (Joseph & Linley, 2006).

Whilst Joseph and Linley (2006) advocate positive psychology going further in disentangling itself from the disease model, its emergence has brought about a renewed impetus to promote psychological research and practice that focuses on positive functioning (Kilbourne *et al.*, 2018). Along with this, there has been a call for outcome measures consistent with, and rooted in, not just positive psychology, but humanistic psychology, discarding the disease model paradigm, and instead focusing on growth and potentiality (Joseph & Murphy, 2013, Joseph & Wood, 2010; Patterson & Joseph, 2006; 2007).

1.3.6. Well-being measures

Patterson and Joseph (2006; 2007) argue that there have been limited options for services and practitioners wishing to monitor outcomes, but who adhere to approaches that do not align themselves with the disorder-focused or medicalised views of human suffering (Patterson & Joseph, 2006; 2007). However, in the wake of the positive psychology movement, several outcome measures that emphasise well-being and optimal functioning have been developed. Below I briefly discuss two of these, the Warwick-Edinburgh Mental Well-being Scale (WEMWBS; Tennant *et al.*, 2007) and the Strathclyde Inventory (SI; Freire, 2007), using them as examples of positive psychology and humanistic psychology measures, respectively.

1.3.6.1. The WEMWBS

The WEMWBS (Tennant *et al.*, 2007) has been widely employed in the UK (Fat, Scholes, Boniface, Mindell, & Stewart-Brown, 2017), and is the only well-being measure included in the national health surveys for England and Scotland (NatCen for Social Research, 2018). Population

norms for the WEMWBS have also been published for France, Spain, Italy and Northern Ireland, and it has been translated into 11 languages (Fat, *et al.*, 2017; Warwick Medical School, 2018). It is a measure that focuses on mental well-being rather than distress and pathology. Prior to the WEMWBS, several widely used scales focused on psychological well-being, such as the Positive and Negative Affects Schedule (PANAS; Watson, Clark, & Tellegen, 1988), focusing on affective-emotional facets of well-being, the Satisfaction With Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985) aiming to measure cognitive-evaluative aspects, and the Scale of Psychological Well-being (SPWB; Ryff, 1989; Ryff & Keyes, 1995), concentrating on eudaemonic well-being and psychological functioning. The WEMWBS (Tennant *et al.*, 2007) builds on all of the above, aiming to cover a broader conception of mental well-being by measuring affective-emotional and cognitive-evaluative dimensions, as well as psychological functioning. Whereas the scales listed above all included both positively and negatively worded items, the WEMWBS is purely focused on the positive, and is short enough (14 items) to be used as an outcome measure (Tennant *et al.*, 2007). It generally has excellent psychometric properties, and had good internal reliability (alpha) in both student ($\alpha = .89, n = 348$) and population samples ($\alpha = .91, n = 1749$; Tennant *et al.*, 2007).

1.3.6.2. The Strathclyde Inventory (SI)

The SI (Freire, 2007) is another scale that aims to measure positive functioning rather than distress or pathology. Patterson and Joseph (2006) argued that there is a need for outcome measures that are rooted in psychotherapeutic theory, and unlike the WEMWBS (Tennant *et al.*, 2007), the SI (Freire, 2007) meets this call, fitting into the narrower category of humanistic measures rather than just a positive psychology measure. It is based on person-centred theory and Carl Rogers' (1961a) description of 'the fully functioning person' (p. 183), and was developed to be a measure that is consistent with the concept of outcome in humanistic therapies. On a 5-point

Likert scale, ranging from 0 ('never') to 4 ('all or most of the time'), t asks respondents to rate how often each of the 16 statements have been true for them over the past month. Of the 16 items, seven are reverse coded. The following are example items: 'I have experienced very satisfying personal relationships' (item 5), 'I have found myself "on guard" when relating to others' (item 9) and 'I have hidden some elements of myself behind a mask' (item 13) (Freire, Elliott, & Rodgers, 2012). Since its initial development, it has been revised several times (e.g. Freire, 2007; Freire *et al.*, 2012; Stephen & Elliott, 2017), with the latest version reporting an internal reliability of .91 in a sample of 405 clients receiving person-centred therapy. Its use as an outcome measure is currently being further researched.

1.3.7. Authenticity

The emergence of the field of positive psychology has brought an interest in research looking at authenticity, a central construct within this framework (e.g. Joseph & Wood, 2010; Seligman & Csikszentmihalyi, 2000). Prior to this however, authenticity had been conceptualised by many different psychological and philosophical perspectives.

In Ancient Greece, Aristotle held that self-realisation comes by discovering one's soul, and that this is fundamental to living authentically, which in turn lends purpose to life (Hutchinson, 1995). Since then, notable philosophers such as Descartes, Hume, Kierkegaard, Heidegger and Sartre have all discussed authenticity at length (Bond, Strauss & Wickham, 2018; Kernis & Goldman, 2006; Markus & Nurius, 1986). Sartre (1944) wrote of authenticity as consisting in having a 'true and lucid consciousness of the situation' (p. 65), seeing things as they are, and accepting them along with the concomitant emotions and responsibilities that come from such a true consciousness.

Heidegger (1962) wrote about how we tend to drift away from our uniquely individual experience of being in the world, into simply being a part of 'the they', forgetting ourselves and instead being absorbed into the collective. The immediacy of our experience of being becomes averaged out into the common world of existence around us, and our unique individuality is surrendered to the commonly defined expectations and styles of living in our surroundings (Sherman, 2009). Heidegger stated that in a world of inauthenticity 'everyone is the other, and no one is himself [*sic*]' (Heidegger, 1996, p. 120). He held that authenticity is the reclaiming of ourselves from enmeshment with 'the they', a shift in attention and engagement to glimpse our unique and individual experiences of being (Sherman, 2009).

Building on these philosophical perspectives, the past century has seen many different schools of psychology explore authenticity, such as social (e.g. Deci, 1980; Kernis & Goldman, 2006; Lopez & Rice, 2006), clinical (e.g. Ehlers, Maercker, & Boos, 2000), evolutionary (e.g. Sedikides & Skowronski, 1997), psychodynamic (e.g. Horney, 1951; Winnicott, 1965), developmental (e.g. Harter, Marold, Whitesell, & Cobbs, 1996), existential (e.g. May, 1981), humanistic (e.g. Rogers, 1961a; Maslow, 1968), and finally, positive (e.g. Sheldon, 2004, Ryff & Singer, 2008). Despite the numerous different conceptualisations of the construct, there are some common denominators, such as the extent to which one's emotions, cognitions and behaviours reflect a core-, organismic- or true self, and a non-defensiveness to one's experiences and to evaluative information (e.g. Deci, 1980; Kernis & Goldman, 2006; Rogers, 1959; Winnicott, 1965; for reviews, see Harter, 2002; Sheldon, 2004).

1.3.7.1. State and trait authenticity

One key point of contention regarding the authenticity construct is whether it is best understood as a trait or as a state. The state-trait debate has been frequently discussed in the

personality psychology literature (i.e. Funder, 1991; Nezlek, 2007), with a trait held to be a person's base-rate propensity towards a set of affects, cognitions or behaviours, whereas a state is considered the actual affects, cognitions or behaviours in a specific situation (Endler, Parker, Bagby, & Cox, 1991). Authenticity has historically been conceptualised within the trait framework, and indeed the vast majority of the literature has explored it in this way (i.e. Kernis & Goldman, 2006; Rogers, 1959; Wood *et al.*, 2008), holding it to be a stable characteristic. Within this framework, any change in authenticity would typically be gradual over time, or facilitated by psychological intervention. Empirical research based on this perspective might ask participants to report to what degree they generally feel, think and act in accordance with set criteria supposedly indicative of authenticity (i.e. Goldman & Kernis, 2002; Wood *et al.* 2008).

However, recent years have seen a growing interest in *situationism*, the view that experienced authenticity changes from situation to situation, which aligns with the broader state view (i.e. Lenton, Slabu, Sedikides, & Power, 2013). For example, Fleeson and Wilt (2010) found that authenticity seemed to vary more within than between people in their studies, with participants reporting to feel more authentic in certain situations, regardless of their reported levels of trait authenticity. Similarly, Lenton, Bruder, Slabu, and Sedikides (2013) found that people were eager to seek out experiences of authenticity and to avoid experiences of inauthenticity, regardless of their levels of trait authenticity. Feeling authentic or inauthentic did not seem to be default states, but rather temporary experiential phenomena and they concluded that state and trait authenticity are qualitatively separate. Importantly, these findings (Fleeson & Wilt, 2010; Lenton *et al.*, 2013) do not suggest a state view of authenticity is more appropriate than a trait view, nor that the two are mutually exclusive, but rather that state and trait authenticity are two distinct phenomena.

1.3.8. Authenticity in person-centred psychology

Within the different schools of psychotherapy, authenticity is a much discussed construct, usually seen as a stable trait that can be increased through psychological intervention. Indeed, several prominent models see authenticity as a key concept, fundamental to understanding human distress and well-being, including psychodynamic (e.g. Horney, 1951; Kohut, 1971; Winnicott, 1965), existential (e.g. May, 1981; Yalom, 1980), gestalt (Perls, 1969) and humanistic models (e.g. Rogers, 1959). The person-centred approach within humanistic psychology, in particular, holds authenticity as pivotal, seeing increased authenticity, which Rogers (1959) termed *congruence*, to be the central outcome and aim of psychotherapy, intrinsically linked to subjective and psychological well-being (Rogers, 1959). Despite its centrality in his psychological theories, Rogers did not introduce one, clear and succinct definition of congruence. Instead, he used several terms interchangeably with congruent, such as ‘real’, ‘genuine’, ‘whole’, ‘integrated’ and ‘transparent’ (e.g. Rogers, 1951, 1957, 1959, 1980). However, it is clear from his writings that he saw congruence as the key outcome of successful psychotherapy, clearly delineated in his aetiological theories.

Rogerian (1959) theory holds that incongruence is the aetiological root of all psychopathology. Incongruence stems from there being a mismatch between a person’s *organismic experience* (i.e. the unfiltered experiencing of the person), and what the *self-concept* (i.e. the way the person sees themselves), is allowing into awareness (Cooper, 2013). When awareness does not accurately match the real, or unfiltered experience, filtering out those experiences that do not match the self-concept, the person is in a state of incongruence. It follows then, that congruence is a mirroring between the full range of the individual's experiences, and the experiences they are symbolising in awareness (Wyatt, 2001). Rogers (1980) wrote that congruence is a close matching, “between what is being experienced at a gut level, what is present

in awareness, and what is expressed' (p. 116). He also described nine characteristic directions of development the client experiences through the process of therapy when these conditions are met (Rogers, 1959). Every one of these listed directions are based on congruence: e.g. '6. He [*sic*] experiences fully, in awareness, feelings which have in the past been denied to awareness, or distorted in awareness' (1959, p. 216). He also listed fifteen personality and behavioural outcomes of psychotherapy. The first of these, upon which the others are built, is: '1. The client is more congruent, more open to his [*sic*] experience, less defensive' (1959, p. 218). From this it is clear that Rogers (1959) saw the aim of person-centred psychotherapy as the increase of congruence, which in turn will reduce psychopathology and bring about positive personality and behavioural change.

Barrett-Lennard (1998) reviewed Rogers's writings, aiming to provide a succinct and clearly bounded definition of congruence. His definition follows Rogers's (1959, 1980) work closely, in addition to being based on substantial debate among person-centred practitioners, resulting in a definition with wide consensus in the person-centred field (Wood *et al.*, 2008; Wyatt, 2001). His definition is a tripartite model, with congruence being: 'consistency between the three levels of (A) a person's primary experience, (B) their symbolised awareness, and (C) their outward behaviour and communication' (Barrett-Lennard, 1998, p. 82; see figure 1).

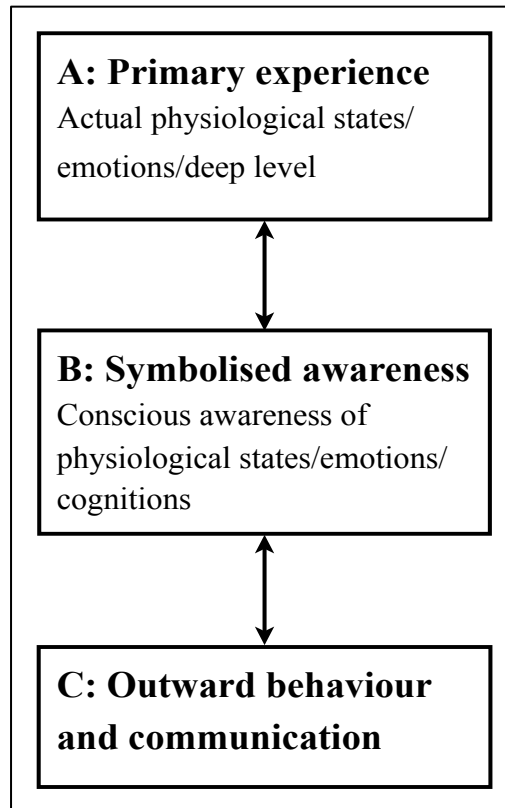


Figure 1. Tripartite model of congruence (Barrett-Lennard, 1998)

1.3.9. Measuring authenticity

Despite the importance placed on authenticity within many psychological and psychotherapeutic perspectives, there has been a relative lack of empirical research looking at this construct (Knoll, Meyer, Kroemer, & Schröder-Abé, 2015). This is in part due to there being few psychometrically adequate measures of authenticity available (Knoll *et al.*, 2015). Sheldon (2004), as well as Peterson and Seligman (2004), argues that authenticity is problematic to operationalise via direct self-report, as one might be unwilling to admit to inauthenticity, one might not be aware of being authentic, and asking participants how authentic or inauthentic they are might simply be too sophisticated a question for some to reflect on. Harter (2002), on the other hand, suggests this relative lack of measures simply reflects the lack of consensus regarding the boundaries and

content of the construct, with so many different perspectives conceptualising authenticity in different ways. Regardless of the underlying reasons, authenticity has been relatively under-researched despite it historically being of great interest to many human sciences, and central in several models of psychotherapy (Harter, 2002; Knoll *et al.*, 2015; Sheldon, 2004).

I conducted a thorough search of the psychology literature with the aim of identifying all available psychometric measures of individual authenticity. I began by searching ISI Web of Knowledge for any psychology or social sciences articles with ‘authenticity’, ‘congruence’ or ‘true or false self’ in their titles, and ‘measure’ or ‘scale’, and ‘validity’ or ‘reliability’ in their topic, using the relevant Boolean operating terms. I assessed the titles of each of the 126 publications yielded, as well as reviews of the authenticity literature. I identified six reviews of the psychological authenticity literature (Harter, 2002; Knoll *et al.*, 2015; Mengers, 2014; Peterson & Seligman, 2004; Sheldon, 2004; White, 2011), and reasoned that together, these would cover any measures of authenticity available before 2015, as that as the most recent review. The literature search identified five measures of authenticity, and these were also covered in the reviews.

Below I briefly discuss the measures of authenticity available, each with different foci and based on a different ways of conceptualising and operationalising authenticity. I then move on to discussing the Authenticity Scale (Wood *et al.*, 2008), the focus of this thesis.

1.3.9.1. The Experienced Authenticity Measure (EAM)

The EAM (Sheldon, Ryan, Rawsthorne, & Ilardi, 1997) was developed to test how the authenticity construct relates to Big-Five personality traits (Costa & McCrae, 1985) within different social contexts, thus conceptualising authenticity as a contextual state rather than a trait. It consists of five items, and participants were asked to answer these for five different social roles. In their sample, reported coefficient alphas ranged from .72 (employee role) to .82 (child role)

depending on the social role. Sheldon *et al.* (1997) found that cross-role variability EAM scores predicted cross-role variability on the Big-Five traits across all social roles, with higher scores on the EAM predicting higher scores on agreeableness, extraversion, conscientiousness and openness to experience, and lower scores on neuroticism. They concluded that authenticity, as operationalised by their scale, is closely linked to well-being.

1.3.9.2. The Authenticity in Relationships Scale (AIRS)

Similarly to the EAM (Sheldon *et al.*, 1997), the AIRS (Lopez & Rice, 2006) is based on a social-contextualist view of authenticity, seeing it not as a stable and invariable trait, but as varying with relationships. This view is consistent with previous research, suggesting authentic behaviour to be more frequent in emotionally significant relationships, such as relationships with family and romantic partners, as opposed to those with acquaintances (DePaulo & Kashy, 1998). The AIRS consists of 24 items, split into two subscales that correlate moderately ($r=.47$, $N=480$): ‘unacceptability of deception’ (UOD; 13 items) and ‘intimate risk taking’ (IRT; 11 items). Lopez and Rice (2006) reported coefficient alphas of .87 and .86, and test-retest stability of .70 and .76, respectively. For both the UOD and IRT subscales, significant positive correlations were reported with the Rosenberg Self-Esteem Scale ($r=.25$, .26; Rosenberg, 1965), and negative correlations with a measure of depression ($r=-.37$, $-.25$; CES-D; Radloff, 1977).

1.3.9.3. The Perception of False Self Scale (POFS)

The PFOS (Weir & Jose, 2010) was developed specifically to be used with adolescents (aged 11-16), measuring the extent to which they perceive themselves as false or inauthentic, and the discrepancy between their public and private selves. It consists of 16 items on a single factor, and the authors reported internal reliability of .88 ($N = 195$; Weir & Jose, 2010). Weir and Jose

(2010) do not specify if they view false self-perceptions as a trait or state, but found that they remained relatively stable over the 10-week test-retest period ($r = .84$). The POFS was also found to correlate at about $r = .65$ with measures of depressive (The Children's Depression Inventory, Kovacs, 1985) and anxious symptoms (The Revised Children's Manifest Anxiety Scale, Reynolds & Richmond, 1997).

1.3.9.4. The Authenticity Inventory (AI)

Kernis and Goldman (2006) subscribe to a trait view of authenticity, as opposed to the authors of both the EAM (Sheldon *et al.*, 1997) and the AIRS (Lopez & Rice, 2006), holding it to be a stable disposition rather than varying with context or relationships. They defined authenticity as 'the unobstructed operation of one's true- or core-self in one's daily enterprise' (p. 294). Building on this definition, as well as a thorough examination of the psychological literature relating to authenticity, they construct multicomponent view of authenticity, as opposed to viewing it as a unitary process. They developed the AI to measure four interrelated, but distinct, components: *awareness* (12-items), *unbiased processing* (10 items), *behavioural* (11 items), and *relational orientation* (12 items) (Goldman & Kernis, 2004). *Awareness* refers to a comprehensive awareness of yourself: the knowledge of, and trust in, your motives, emotions, desires, and cognitions. *Unbiased processing* is the absence of denial or distortion of deep experiences and emotions, an openness to evaluative information, and acceptance of your strengths and weaknesses. *Behavioural* refers to acting in accordance with your values, beliefs and needs, rather than merely trying to meet external expectations; and *relational orientation* is the drive to openness and honesty in close relationships, allowing those close to you to see the authentic and real you (Kernis & Goldman, 2006). The final Authenticity Inventory (AI-3; Goldman & Kernis, 2004) has 45 items, and the authors reported a coefficient alpha of .90 for the full scale, with scale alphas

ranging from .64 (unbiased processing) to .80 (behavioural) for the four subscales. They also reported 4-week test-retest reliability of .87 for the total scale (Kernis & Goldman, 2006). Higher scores on the AI-3 were related to higher scores on measures of self-actualisation tendencies (Jones & Crandall, 1986; $r=.61$), mindfulness (MAAS; Brown & Ryan, 2003; $r=.49$) and lower scores of psychological distress (Cohen, Kamarack, & Mermelstein, 1983; $r=.30$). However, some later research has failed to replicate the factor structure of the AI-3, with a study finding that several items did not load onto any one factor, casting doubt on the validity of the measure (White, 2011).

Recently, work has been done to develop a shorter and more psychometrically sound version of the AI-3, the Kernis-Goldman Authenticity Inventory - short form (KGAI-SF; Bond, *et al.*, 2018). Thirteen of the original 45 items were removed, being judged semantically ambiguous or redundant, and confirmatory factor analysis revealed that a further 12 items were unreliable or psychometrically redundant. The final KGAI-SF consists of 20 items, and Bond *et al.*, (2018) reported internal reliability of .87 ($N = 1252$) for the full scale.

1.3.10. The Authenticity Scale

The above measures all approach the conceptualisation of authenticity from different directions. However, none of them are based in psychotherapeutic theory, nor were they developed as potential outcome measures. The Authenticity Scale (AS; Wood *et al.*, 2008) on the other hand, was developed specifically to meet the need for a short outcome measure for counselling psychology practice and research, grounded in psychotherapeutic theory.

Whereas the AI was developed out of a broad reading of the literature on authenticity, pulling together its four factors from a variety of different perspectives, the AS is a measure based on theory, with items developed specifically to operationalise the person-centred construct, congruence (Wood *et al.*, 2008). The overlap with the SI (Freire, 2007) however, is much larger

than with the AI, or with any of the other authenticity measures, though the SI does measure a broader construct than the AS, as it is based on Rogers' (1959) descriptions of the fully functioning person. The fully functioning person is what Rogers believed the client would be closer to after successful therapy, and increased congruence was thought to be the driving factor in this process. Thus, the AS and SI set out to measure related, yet distinct constructs: the SI measures how closely the person fits with how Rogers' believed a client would be after successful therapy, whereas the AS aims to measure the antecedent of this, the causal factor in become fully functioning.

Specifically, Wood *et al.* (2008) based the AS on Barrett-Lennard's (1998) tripartite model of congruence, derived from Rogers's works (1959, 1961a, 1961b) (see figure 1). They defined authenticity as consistency between the three levels of the tripartite model (Barrett-Lennard, 1998), with *self-alienation* (1) consistency between primary experience (A) and symbolised awareness (B), and *authentic living* (2) consistency between symbolised awareness (B) and outward behaviour and communication (C) (see figure 2). These two facets of authenticity are also present in the work of Schmid (2001, 2005), where he refers to self-alienation as 'openness to oneself' (p. 216) and authentic living as 'transparency' (p. 216), and in the work of Lietaer (1993).

In addition to self-alienation and authentic living, Wood *et al.* (2008) added a third process to their model: *accepting external influences* (3). This aspect of authenticity is the extent to which the person introjects the views of others, and believes he or she must conform to their expectations and demands (Wood *et al.*, 2008). Rogers (1959, 1961a, 1961b) described the fully functioning, congruent person, as having an *internal locus of evaluation*, that is, self-directivity and moving away from needing to please others and meet their expectations (Bohart, 2013). He also included this in his list of psychotherapy outcomes in personality and behaviour, holding that this was intrinsically tied to congruence and a key outcome of psychotherapy (Rogers, 1959). Figure 2 illustrates the model of authenticity upon which Wood *et al.* (2008) based the AS.

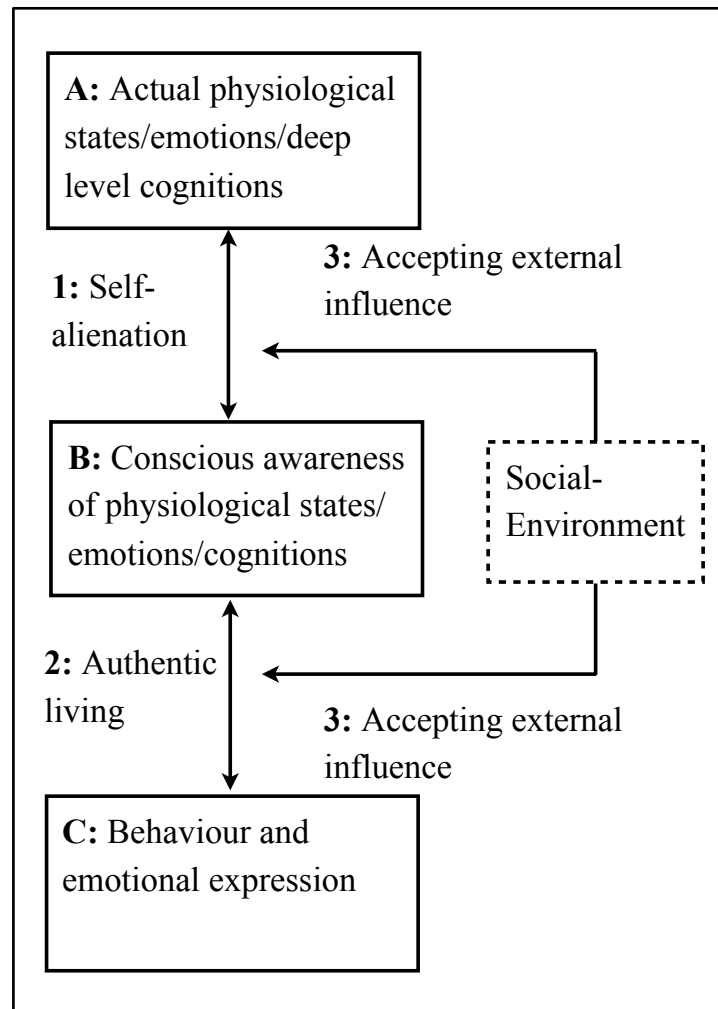


Figure 2. The conceptual basis for the Authenticity Scale (Wood *et al.*, 2008)

1.3.10.1. The Authenticity Scale as an outcome measure

As Rogerian theory sees congruence as the ultimate aim of psychotherapy (Rogers, 1959), Joseph and Murphy (2013) suggested that using a measure of congruence, such as the AS, would be an appropriate way in which to evaluate therapy. The AS was developed with this aim, as well to help explore the link between authenticity and well-being (Wood *et al.*, 2008), aiming to provide counselling psychology and the psychotherapy community with a well-being orientated measure that was conceptually based in psychotherapeutic theory, thus evaluating therapy on its own premises. It was limited to 12 items specifically to make it easier to implement in counselling

settings (Wood *et al.*, 2008) and, indeed, White (2011) concluded, after examining the use of the AS in a North American sample, that it could be particularly beneficial as a measure in the field of counselling psychology because of its length and robust psychometric properties. Though it is based on person-centred theory and humanistic psychology, Wood *et al.* (2008) held it to be an appropriate outcome measure for psychotherapeutic approaches beyond just humanistic ones, as several of the most prominent models highlight authenticity as a key therapeutic outcome (i.e. Horney, 1951; Kohut, 1971; May, 1981; Winnicott, 1965, Yalom, 1980), and their conceptualisations of authenticity overlap.

1.3.10.2. The psychometric properties of the Authenticity Scale

The 12 items of the AS make up three subscales corresponding to the three processes of the conceptualisation of authenticity adopted by Wood *et al.* (2008): *authentic living* (AL), *self-alienation* (SA), and *accepting external influences* (AEI). High scores on AL are indicative of high authenticity, whereas the SA and AEI are reverse scored, with high scores on these two indicative of low authenticity. The factor structure was first confirmed by exploratory factor analysis (EFA) on the initial item pool, before multigroup confirmatory factor analysis (CFA) was conducted on the final items, providing clear support for the three-factor model, and the three subscales were all moderately intercorrelated in three separate samples ($n = 180, 158, 210$). These correlations were as expected by the authors, and provided further support for the factor structure of the AS, suggesting the subscales were interrelated, but measured distinct and separate constructs (Wood *et al.*, 2008). The authors reported coefficient alphas of .70 - .82 for AL, .82 - .84 for SA, and .77 - .84 for AEI. They also reported four-week test-retest reliabilities at $r = .78, .79$ and $.81$ for the subscales, respectively (Wood *et al.*, 2008).

Socially desirable responding did not appear to be an issue for the AS, with low and non-significant correlations with the two subscales of the Balanced Inventory of Desirable Responding scale (Paulhus, 1984) for both the full AS and each of the three subscales (Wood *et al.*, 2008). Since its development, the AS has been used in numerous studies, and has generally been found to have excellent psychometric properties. I will explore these in chapter 2.

1.4. Conclusion

In this chapter I have reviewed the development of outcome monitoring in counselling psychology, beginning with the Eysenck study (1952) and continuing to the recent focus on evidence-based practice and the current climate of high accountability for psychological therapies (BPS, 2017; Evidence-Based Medicine Working Group, 1992; HCPC, 2015; Tarescavage & Ben-Porath, 2014). After reviewing a selection of commonly used outcome measures, I discussed the emergence of the positive psychology perspective, and the accompanying call for outcome measures that measure well-being and optimal functioning rather than illness, dysfunction and suffering (Levitt *et al.*, 2005; Patterson & Joseph, 2006). Authenticity is intricately interlinked with well-being, a key construct not only within positive psychology but also within several major models of psychotherapy, such as psychodynamic, existential and humanistic (May, 1981; Rogers, 1959; Winnicott, 1965). Authenticity has been conceptualised and operationalised in several different ways, but the Authenticity Scale (Wood *et al.*, 2008) is the only measure specifically developed to be an outcome measure for psychotherapy. Based on the person-centred definition of congruence (Barrett-Lennard, 1998) the AS is a measure consistent with psychotherapeutic theory, thus measuring the effectiveness of therapy on its own premises. Below I outline the aims of this thesis, building on the rationale provided in this chapter.

1.5. A priori aims

This thesis has two overall aims, addressed over two studies:

- i. Investigate the psychometric properties of the AS.** This includes looking at its internal reliability, and its concurrent, construct and factorial validity.
- ii. Investigate how the AS functions as an outcome measure.** This includes looking at its acceptability, in the form of its Therapy Questionnaire Helpfulness Survey (TQHS; Di Malta, Pauli & Cooper, unpublished) scores, and completion rates, its score distributions, and pivotally, its sensitivity to change.

STUDY 1

Chapter 2. Systematic review of the psychometric properties of the Authenticity Scale, with a narrative synthesis and meta-analysis

2.1. Introduction

As noted in chapter 1, the Authenticity Scale (AS; Wood *et al.*, 2008) was developed as a positive psychology measure, one not based on illness ideology measuring the extent of suffering, but rather well-being and healthy functioning. It aims to measure authenticity, based primarily on the construct ‘congruence’ from the field of person-centred psychology as defined by Barrett-Lennard (1998; Rogers, 1959). Ten years after its development, no publication has systematically reviewed the literature pertaining to the Authenticity Scale, despite a high level of interest in authenticity as a construct (e.g. Harter, 2002; Robinson, Lopez, Ramos, & Nartova-Bochaver, 2012), and a considerable number of studies employing the AS in various ways. The present review aims to investigate the psychometric properties of the AS, collating data from all publications that have applied it in some way, as well as to investigate whether it has been used as an outcome measure for psychotherapy, or in clinical populations. The review is based on a systematic literature search and consists of two parts, (a) a narrative synthesis of the AS literature, exploring the concurrent validity of the AS, and whether it has been employed as an outcome measure, or with clinical populations, and (b) a meta-analysis of coefficient alphas, as well as of mean AS scores and subscale intercorrelations.

2.2. Methods

2.2.1. Aims of the systematic review and research questions

This systematic review aims to address the first aim of the thesis. Thus, its overall aim is to investigate the psychometric properties of the AS in the published literature. This aim is split into four subsidiary aims:

1. **Explore the concurrent validity of the AS in the literature.** This involves summarising what relevant measures the AS has been correlated with in the literature, and the nature of these correlations.
2. **Explore the internal reliability of the AS in the literature.** This involves meta-analysing coefficient alphas for AS scores, and for its subscales.
3. **Explore construct validity of the AS in the literature.** This involves meta-analysing the subscale intercorrelation coefficients between the AS and its subscales, and considering them in light of their coefficient alphas.
4. **Explore the mean AS scores in the literature.** This involves meta-analysing the mean AS scores, and those of its subscales.

Finally, the second aim of the thesis was to investigate how the AS functions as an outcome measure. Thus, I included a fifth aim for the systematic review:

5. **Explore the use of the AS as an outcome measure, and with clinical samples.** This involved specifically reviewing the psychometrics of the AS from any studies that had used it with clinical samples or as an outcome measure.

These five aims for Study 1 will be addressed by the following research questions:

1. Does the AS correlate in expected ways with other psychometric measures?
2. How internally reliable is the AS in the samples from published studies, and how much between-study variability is there in reliability coefficients?
3. Do the subscales of the AS inter-correlate in expected ways in published studies?
4. What are the mean AS scores in the samples from published studies, and how much between-study variability is there in mean scores?
5. Has the AS been used as an outcome measure, or with clinical samples, and if so, what were its psychometric properties in these studies?

2.2.2. Methods of the systematic review

To address the above aims, I decided to split this systematic review in two parts. To explore the concurrent validity of the AS, I opted for a narrative synthesis, as I expected there to be relatively little overlap across publications in what measures the AS had been correlated with. Thus, collating correlation coefficients seemed inappropriate, whereas a narrative synthesis would allow me to group measures together and thoroughly discuss the convergent and divergent validity of the AS. Finally, I also planned to address the fifth aim narratively, allowing me an in-depth exploration of any studies that had applied the AS as an outcome measure, or with clinical samples.

I also reasoned the majority of studies would report coefficient alphas and mean scores of the AS, as well as its subscale intercorrelations, and thus it seemed natural to address aims two, three, and four, in a meta-analysis, allowing me to collate coefficients and means, and explore the between-study variance.

2.2.2.1. Registration of the systematic review

This systematic review is registered in the International Prospective Register of Systematic Reviews (PROSPERO; www.crd.york.ac.uk), with registration number: CRD42018116038. Where relevant, it follows the Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines (PRISMA; Moher *et al.*, 2009), a 27 item checklist and flow diagram for researchers conducting meta-analyses and systematic literature reviews. Though several of the points are not relevant for a thesis, such as a specific abstract, I registered this review with PROSPERO as I intend to rewrite it for a potential publication.

2.2.2.2. Eligibility criteria

The systematic literature search aimed to identify every publication that included data on the reliability or validity of the AS, as well as AS scores. Thus, the eligibility criteria for inclusion was simply that the publication reported AS data, and searches were not refined in any way (e.g. by field, topic, publication type etc). I decided that translations of the AS would be eligible, as these are still aiming to measure the same constructs, though the psychometric properties may be different. I also decided that adaptations of the AS that had either reworded items, removed items, or added items, would not be eligible, as this would mean the measure was no longer the Authenticity Scale as intended (Wood *et al.*, 2008).

2.2.2.3. Data sources and search strategy

All electronic information databases from two comprehensive research platforms were searched to identify relevant publications. These were ISI Web of Knowledge (Web of Science Core Collection, BIOSIS Citation Index, Chinese Science Citation Database, SciELO Citation Index) and EBSCOhost (PsychARTICLES, Academic Search Premier, PsychINFO, Business

Source Premier, Education Research Complete, SPORTDiscuss, PsychBOOKS, Communication & Mass Media Complete, ERIC). The most recent search of the ISI Web of Knowledge and EBSCOhost databases was on October 31st, 2018.

As the aim of the search was to identify any research with AS data, I reasoned that the common denominator for all relevant publications would be the citing of Wood *et al.* (2008). In ISI Web of Knowledge it was possible to specifically search for articles citing Wood *et al.* (2008). This was not possible with EBSCOhost, but I reasoned that all relevant publications would include the title of the AS paper in their reference list, and thus I searched ‘all text’ for ‘The Authentic Personality: A Theoretical and Empirical Conceptualization and the Development of the Authenticity Scale’ (Wood *et al.*, 2008). All results were downloaded into the Mendeley Desktop software (Version 1.19.2; Mendeley Ltd, 2018) to be assessed for eligibility.

2.2.2.4. Study selection

The search strategy yielded 272 publications, 208 from ISI Web of Knowledge databases, and 64 from EBSCOhost databases. Of these, 30 were duplicates, leaving 242 publications for review.

I reasoned that just screening articles by abstract would not be sufficient to determine whether they contained data on the AS, so I assessed all 242 publications for eligibility, which meant searching the full-text versions of each publication for empirical AS data. Through this process, 157 studies were excluded because they did not report any AS data, and a further 23 were excluded for having changed the AS in some way, leaving 62 publications relevant to this study. After reviewing these, I excluded a further two for reusing samples from previous publications. This left 60 studies to be included in the meta-analysis (see figure 3).

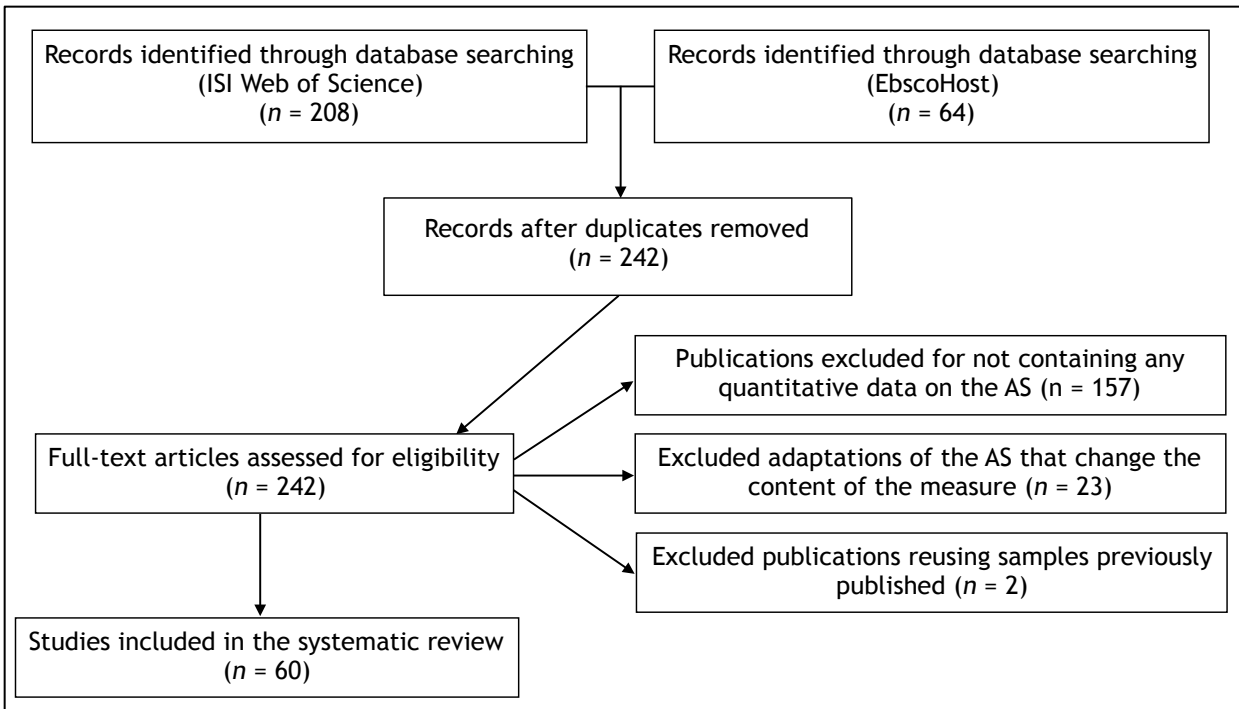


Figure 3. Flow diagram detailing the different phases of the literature search (Mohler et al., 2009).

2.2.3. Methods of the narrative synthesis

The aim of the narrative synthesis was to explore the concurrent validity of the AS, as well as its use as an outcome measure and in clinical samples.

2.2.3.1. Construct validity

As a latent construct, such as authenticity, cannot be directly observed, a psychometric scale aims to operationalise the construct, attempting to make it measurable through a series of items. The measure's validity is how successfully it operationalises the construct, essentially whether the variance in the measure's items is caused by the construct it aims to measure (DeVellis, 2012). There is of course, no way to fully ascertain whether a measure has successfully operationalised the construct it seeks to measure, and thus, validity cannot actually be established.

It follows that there is no single statistical test of validity, but rather it is assessed in different ways that together build a case for the measure's validity.

Generally, construct validity refers to how well the scale measures what it purports to measure, based on scientific analyses and reasoning (Markus & Lin, 2010). Concurrent validity is one aspect of construct validity and refers to whether measure's scores correlate in expected ways with another measure's, administered at the same time. If they do, it supports the notion that the measure is tapping the construct it aims to measure, thereby providing validity (Vogt, 2005). Concurrent validity in turn, is often split into convergent and divergent validity.

To explore the concurrent validity of the AS in Study 1, I reviewed all 60 publications found by the systematic literature search, and identified the ones that reported correlations between the AS and another measure. As there were too many measures to discuss each one, I grouped them by the constructs they aimed to measure, and discussed how the AS relates to the groups of measure most represented in the AS literature, focusing on the ones whether the conceptual links to authenticity are clearest, such as well-being, emotional distress, and other measures of authenticity.

2.2.4. Methods of the meta-analysis

The aim of the meta-analysis was to provide a statistical summary of the internal reliability, subscale intercorrelations, and means, of the AS, from the 60 identified studies. Thus, data was collated into three main parts, (1) an exploration of *reliability generalisation* (Vacha-Haase, 1998), that is, a meta-analysis of Cronbach's alpha coefficients; (2) a meta-analysis of subscale intercorrelations; and (3) a meta-analysis of estimated mean scores.

2.2.4.1. Data analysis

A key obstacle I faced with the meta-analytic part of this study, was that the statistical software I was familiar with was not able to perform the meta-analyses I wished to conduct. After discussions with my supervisory team, we decided the ‘R’ software (Version 3.5.2; R Core Team, 2018) with the Metafor package (Viechtbauer, 2010) would be most appropriate and powerful for this analysis. However, I was not sufficiently familiar with the ‘R’ software to perform the analyses I wanted, nor did I have the time needed to become proficient. Thus, after I had conducted the systematic literature search, decided upon the analyses I wished to conduct, and compiled all the data in a format appropriate for ‘R’, the actual analysis was carried out by Professor Chris Evans in consultation with myself, the main author. Importantly, I understand all the analyses and statistics we have used, and I take responsibility for the analysis. Moreover, I have interpreted the output from ‘R’, and written the entirety of this chapter alone, as well as the thesis.

In the following sections, I briefly outline and explain the statistical methods used.

2.2.4.2. Random-effects model

A meta-analysis essentially combines the effects from several studies into one, estimated combined effect. However, as the precision of the studies varies, such collation needs to assign more weight to the more precise studies, that is, the studies with smaller standard error (*SE*). The way in which such weighting is assigned, and the combined effect calculated, depends on whether a fixed-effects model, or a random-effects model is used (Borenstein, Hedges, Higgins, & Rothstein, 2010). A *fixed-effects model* essentially assumes that there is one true effect which is the same for every included study, and thus the combined effect is an estimate of this true effect, and the larger the sample, the greater the precision of this estimate. A *random-effects model* on the other hand, assumes that the between-studies heterogeneity in effects is due to real differences

between the studies and not just sampling variability (Riley, Higgins, & Deeks, 2011). Thus, the aim is not to estimate one true effect, but rather to calculate the mean of a distribution of effects. Moreover, as each study provides information about the effect in a different population, it is important that even an imprecise study, with large *SE*, contributes to the combined estimate (Borenstein *et al.*, 2011).

2.2.4.3. Calculating the combined effect under a random-effects model

Under a fixed effects model, each study is simply weighted by the inverse of its within-study variance, and then each effect is combined to estimate the true effect. The weighting of each study under a random-effects model is also calculated by the inverse of its variance, but because the effect varies between studies, the variance in this model includes both the within-study variance and the between-study variance (Borenstein *et al.*, 2011). Thus, there is both within-study, and between-study sampling error to take into account when assigning weights to the studies, which means precision of the weighted combined effect depends both on the sample size of each included study, as well as the number of included studies (Borenstein *et al.*, 2011).

This meta-analysis aimed to combine the coefficient alphas, means and intercorrelations of the AS. These properties are not true population effects. For example, it would be unreasonable to expect there to be one true AS score for humanity as a whole, and that each sample is a more or less precise estimate of this true score. Thus, I expected the psychometrics of the AS to vary systematically between-studies, likely relating to variables such as culture, age, language and specific populations, and it followed that it was appropriate to use a random-effects model, which was chosen a priori.

2.2.4.4. Heterogeneity

As discussed above, I expected a large degree of between-study heterogeneity in this meta-analysis. This heterogeneity was assessed in two ways: using (a) the Cochran's Q-test (1954) and (b) the I^2 statistic (Higgins & Thompson, 2002).

Cochran's Q-test (1954) is a chi-squared test, and the classical measure of between-study heterogeneity used in meta-analyses. It assesses whether the observed differences in effects being investigated are compatible with chance alone, by essentially looking at the differences between the observed effect from each study, and the overall combined effect, weighting each study by its inverse variance as discussed above (Huedo-Medina *et al.*, 2006). As I expected the coefficient alphas, means and correlation coefficients to vary considerably between studies, I expected the Cochran's Q-test to be significant, indicating significant between-study heterogeneity.

One weakness with Cochran's Q-test however, is that it does not give the magnitude of heterogeneity, just whether heterogeneity is present or not (Huedo-Medina *et al.*, 2006). The I^2 statistic (Higgins & Thompson, 2002) on the other hand, quantifies the degree of heterogeneity between studies, calculating the percentage of the total variability that is due to between-study heterogeneity (Huedo-Medina *et al.*, 2006). Thus, if $I^2 = 0\%$, it means there was no between-study variance, and that all variability in effects is within-studies. It follows that I expected a high I^2 percentage.

2.2.4.5. Reliability

The reliability of a measure is 'the proportion of variance attributable to the true score of the latent variable' (DeVellis, 2012, p.27). In other words, a reliable measure is one with relatively little random error embedded in the scores. The most widely used index of reliability is coefficient α (alpha; Cronbach, 1951), a coefficient ranging from 0 to 1 assessing the internal consistency of

a measure (Streiner, 2003). It assesses the internal consistency of the items, that is, the extent to which they covary. If the items covary highly, it is likely they are all tapping into some shared issue, measuring the same latent variable. For coefficient alpha, there are cutoff values available meant to help the reader quickly assess the reliability of the measure (e.g. Nunally & Bernstein, 1994). For example, Crutzen and Knutsche (2013) use $\alpha \geq .9$ as 'excellent', $\alpha \geq .8$ as good, $\alpha \geq .7$ as acceptable, and $\alpha \geq .6$ as questionable. However, despite this being convenient, Crutzen and Peters (2017) hold that what is considered acceptable coefficient alpha levels is a matter of professional judgment, and indeed a question of pragmatic judgement, as the required coefficient alpha level depends on what the measure is intended for.

It is important to note that reliability is a function of a measure's *scores*, not of the measure itself, and is therefore a property of the sample. Thus, a measure itself is not reliable or unreliable, but rather reliability needs to be assessed for every sample, though two similar samples will likely yield similar coefficient alpha scores on a measure (Streiner, 2003).

In this study, coefficient alphas for the AS were collected from all publications reporting them, and combined in a reliability generalisation.

2.2.4.6. Construct validity

The AS has a clearly defined design model: it is made up of three distinct scales, that are correlated but oblique, and that together form the whole (Wood *et al.*, 2008). Thus, another way to assess its construct validity is to examine whether data from the publications included in this meta-analysis is congruent with this design model. I did this by looking at the intercorrelations of the three subscales, AL, AEI and SA. Based on the design model of the AS, I expected them to correlate moderately with each other, at around $r = .3$, and to correlate highly with the full scale, at around $r = .7$. Importantly, such correlation coefficients on their own would only attest to these

scales being interrelated, not distinct. If however, they also had high coefficient alphas, it would attest to each measuring something distinct. Thus, if the subscale intercorrelations are to support the construct validity of the AS, they must be considered in light of their coefficient alphas.

In this study, correlation coefficients for the subscale intercorrelations with each other and the overall AS were collected from all publications reporting them, and combined.

2.2.4.7. Publication bias

Publication bias refers to the association between a study's findings and the probability that it will be published (Sterne & Harbord, 2004). The funnel plot is a common way to graphically assess for publication bias in meta-analyses, with study effect on the x-axis, in this case coefficient alpha or AS score means, and an inverse estimate of study size or precision on the y-axis, in this case *SE*. Studies with larger samples will have greater precision (lower *SE*) and will gather high around the combined effect estimate, whilst smaller studies with less precision will scatter lower on the plot. Due to random error we can expect this scatter to be on either side of the overall estimate (Egger, Davey Smith, Schneider, & Minder, 1997). If, however, there is a publication bias, that is, some studies with low coefficient alphas or means may have remained unpublished, this would lead to an asymmetrical funnel plot, missing studies to the left of the overall estimate. This, in turn, would lead to an overestimation of the overall coefficient alpha or mean (Egger *et al.*, 1997).

One way in which to statistically assess for publication bias is testing for funnel plot asymmetry (Egger *et al.*, 1997), and the Begg-Mazumdar correlation test (1994) is one of the ways in which to do this, though Egger *et al.* (1997) make it clear that a visual inspection of the plot is also important. *The Begg-Mazumdar test* (1994) is a Kendall rank correlation between the values on the x-axis, in this case coefficient alpha or AS means, and their *SE*. In the absence of publication

bias, the funnel plot is typically symmetrical and the correlation should be 0 (Egger *et al.*, 1997). A positive correlation however, indicates the plot is asymmetrical with more published studies with higher estimates than the combined estimate (to the right of the funnel plot), whereas a negative correlation indicates asymmetry with more published studies with lower estimates than the overall estimate (to the left of the plot). It follows that a positive correlation is more likely to indicate true publication bias than a negative one.

Funnel plots for both the reliability generalisation and the meta-analysis of means were used to detect potential publication bias, and I used the Begg-Mazumdar correlation test (Viechtbauer, 2010) to test for funnel plot asymmetry.

2.2.4.8. Percentage of maximum possible (POMP) scores

To assess how close to the maximum possible score the AS and its subscales were, I calculated POMP scores. POMP is a standardisation method for scores, giving a percentage of how the raw score compares to the upper and lower limits of the scale. Cohen, Cohen, Aiken, and West (1999) argue that standardising scores into POMP scores is useful in that it put scores into an easily understandable metric, conveying immediate meaning to the reader, and allowing easy comparison between scales using different scoring systems. For example, the AS has a maximum score of 84, whereas each of its subscales have maximum scores of 28, which can make it confusing when trying to interpret scores. By converting scores to POMP however, an AS score of 67 and a score of 18 on the AL subscale, POMP scores of 83% and 58% respectively, can be more easily understood and interpreted.

POMP is calculated by subtracting the minimum score from the raw score, dividing by the maximum score less the minimum score, and multiplying by 100 to get a percentage (Cohen *et al.*, 1999).

2.2.4.9. Dealing with erroneous data

In several cases there were clear and obvious issues with the reported data. I encountered two instances where two papers reported identical means and standard deviations (*SD*), and upon contacting the first authors, I found that they had reused datasets from previous papers. I removed the duplicate datasets from the meta-analysis.

In two instances, studies presented mean scores for both the full AS and its subscales, but the subscale scores did not add up to the full AS score. I emailed the first authors of both studies for clarification. I got a response from one author, who had used an alternative method for adding up the subscales, which allowed me to correct the issue. Seven studies in all had either misreported means, or added means together erroneously, reducing the power of the meta-analysis and potentially inflating the between-samples heterogeneity. I corrected those cases where possible, or removed them from the analysis. In consultation with my supervisor, Professor Chris Evans, I removed two further studies I suspected had wrongly added together raw scores, though because one only reported coefficient alpha, and the other only reported an overall AS mean, it was impossible to fully ascertain what had been done.

2.3. Results

2.3.1. Description of the studies and samples

The systematic review collated data from 85 samples across 60 studies, with sample sizes ranging from $n = 11$ to $n = 1005$. The average sample size was $n = 267$, and the combined sample size was $N = 23,043$. The weighted mean age of participants across all samples was 26.3 years with ages ranging from 15 to 90. Table 1 gives an overview of all the publications included in this systematic review, with sample sizes, sample descriptions, country the study took place in, as well

as an overview of the psychometric data relevant to this study. Thirty-two of the 60 studies used student samples, which was the largest between-samples common-denominator. I classified 14 of the 28 as opportunity samples, as these were highly heterogenous, and difficult to describe succinctly. Participants in these studies were typically recruited by flyers, advertisements, or on websites.

Of the publications included, 53 reported coefficient alphas, 40 reported mean scores, 22 reported subscale intercorrelations, and 37 reported correlations between the AS and other measures that could provide concurrent validity.

Table 1

Overview of all publications included in the systematic review

<i>Reference</i>	<i>N</i>	<i>Sample description</i>	<i>Country</i>	<i>Included data</i>			
				<i>α</i>	<i>Means</i>	<i>Intercorrelations</i>	<i>Concurrent validity</i>
Akfırat et al (2016)	317	University students	Turkey	✓	✓	-	-
Akin & Akin (2014a)	329	Undergraduates	Turkey	-	✓	✓	-
Akin & Akin (2014b)	382	Undergraduates	Turkey	-	✓	✓	-
Akin & Akin (2014c)	366	Undergraduates	Turkey	✓	✓	✓	✓
Barnett & Deutsch (2016)	813	Psychology undergraduates	US	✓	✓	-	-
Boyras & Kuhl (2015)	619	Online opportunity samples*	US	✓	✓	-	✓
Boyras et al. (2014)	232	Students	US	✓	✓	-	✓
Casale et al. (2018)	274	Undergraduates	Italy	-	✓	✓	-
Chiaburu et al. (2010)	55	Professionals on development programmes	US	✓	✓	-	-
Counted & Moustafa (2017)	100	Christian adolescents	South Africa	✓	✓	-	✓
Datu & Reyes (2015)	375	Undergraduates	The Philippines	✓	✓	-	✓
Davis et al (2015)	225	Online opportunity samples	US	✓	✓	-	✓
Di Fabio & Gori (2016)	261	Students and professionals	Italy	✓	-	-	-

Di Fabio & Kenny (2016)	184	Students	Italy	✓	-	-	-
Di Fabio & Palazzeschi (2015)	168	Secondary school pupils	Italy	✓	✓	-	✓
Gil-Or et al. (2015)	258	Undergraduates	US	✓	✓	-	✓
Goldner (2016)	200	Psychotherapists	Israel	✓	✓	-	✓
Grégoire et al. (2014)	625	Students	France	✓	✓	✓	✓
Grijak (2017)	706	University students	Serbia	✓	✓	✓	✓
Guèvremont & Grohmann (2016)	105	Opportunity sample of consumers	US	✓	-	-	-
Güngör et al. (2014)	335	University students	Japan, Turkey	✓	-	-	-
Keng (2017)	23	Student counsellors	Singapore	✓	-	-	-
Kifer & Heller (2013)	351	Opportunity samples	n/a	✓	-	-	-
Lenton et al. (2013)	238	Psychology undergraduates	Scotland	✓	-	-	-
Lenton et al. (2016)	354	Students	Scotland	✓	✓	✓	-
Liang (2017)	90	Professionals	China	✓	✓	-	-
Litman et al. (2016)	1005	Opportunity samples	UK	✓	✓	-	-
Lopez et al. (2015)	100	Undergraduates	US	✓	-	✓	✓
Maltby et al. (2012)	1286	Opportunity samples	UK	✓	-	-	-
McCormick et al (2015)	102	Undergraduates	US	✓	✓	-	✓
Pillow et al (2017)	629	Undergraduates	US	✓	✓	-	✓
Pinto et al. (2011)	554	Community centre attendees	UK	✓	✓	✓	✓
Pinto et al. (2012)	62	Students	UK	✓	✓	✓	✓
Poncy et al. (2018)	191	Students	US	✓	-	-	-
Proctor et al. (2016)	636	Secondary school pupils	UK	-	✓	✓	✓
Riggle et al. (2014)	272	Lesbian/gay/bisexual participants	US	✓	-	-	✓
Robinson et al. (2013)	628	Opportunity samples	Russia, US, UK	✓	✓	-	✓
Robinson et al. (2014)	553	Undergraduates and opportunity samples	UK	✓	-	-	-
Robinson et al. (2015)	737	Students and professionals	UK, Iran, China	✓	-	-	-
Robinson et al. (2017)	963	Opportunity samples	UK	✓	-	-	-
Rollero (2016)	235	Undergraduates	Italy	✓	✓	✓	✓
Sarıçam (2015)	347	Undergraduates	Turkey	✓	✓	✓	✓
Satici, Kayfis et al. (2013)	303	Undergraduates	Turkey	-	✓	✓	-
Satici, Uysal et al. (2013)	294	Undergraduates	Turkey	-	✓	-	-
Seto & Hicks (2016)	304	Online opportunity samples	US	✓	✓	-	-
Seto et al. (2016)	457	Online opportunity samples	US	✓	✓	✓	✓
Shamsi et al. (2012)	304	Students	Iran	✓	-	-	-

Slabu et al. (2014)	622	University students, opportunity samples	US, India, China, Singapore	✓	✓	-	-
Stevens (2017)	116	Undergraduates	US	✓	✓	✓	✓
Tekin & Satici (2014)	284	Undergraduates	Turkey	-	✓	✓	-
Thomaes et al. (2017)	155	Adolescents	UK	✓	-	-	-
Touré-Tillery & Light (2018)	602	Online opportunity samples	US	✓	-	-	-
Vainio & Daukantaite (2016)	592	Students and educational professionals	Sweden	✓	-	-	-
Vess et al. (2014)	83	Online opportunity samples	US	✓	✓	✓	✓
Vess et al. (2016)	199	Psychology undergraduates	US	✓	-	✓	✓
Wang (2015)	505	Undergraduates	China	✓	✓	-	-
Wang (2016)	191	Adult opportunity samples	China	✓	-	-	-
White & Tracey (2011)	537	Undergraduates	US	✓	✓	✓	-
Wood et al. (2008)	751	Undergraduates and professionals	UK	✓	✓	✓	✓
Zhang et al. (2018)	459	University students	China	✓	✓	✓	-

Note: *N* = sample size across all samples; α = coefficient alpha; *Means* = AS means and/or subscale means; *intercorrelations* = correlations between the AS and its subscales; *Concurrent validity* = measures correlated with the AS

*Opportunity samples refers to a heterogenous sample recruited from multiple places.

2.3.2. Results of the narrative synthesis

2.3.2.1. Concurrent validity

The narrative synthesis aimed to summarise the concurrent validity of the AS, based on 26 publications reporting AS correlations with other psychometric measures. Publications that used a translation of the AS, but did not report internal reliability (coefficient alpha) for their sample, were not included, as there was no way to ensure the translation was adequate.

I combined the measures most frequently correlated with the AS into four groups: (a) well-being, with seven different measures represented; (b) emotional distress, with seven different measures; (c) personality, with four different measures; and (d) authenticity, with three measures of authenticity or closely related constructs. I focused on these four groups as they are all highly relevant to the model the AS is based on, and were expected to help elucidate its concurrent validity. Table 2. Shows the measures most frequently correlated with the AS in the literature.

Table 2*Psychometric measures correlated most frequently with the AS*

<i>Measure</i>	<i>Construct</i>	<i>Measure citation</i>	<i>Number of publications</i>
Satisfaction with life scale (SWLS)		Diener et al. (1985)	7
Positive and negative affect scale (PANAS)	Frequency of negative and positive emotions	Watson et al. (1988)	5
Experiences in close relationships scale (ECR)	Anxious and avoidant attachment styles	Brennan et al. (1998)	5
Big five inventory (BFI)	Big-5 personality traits	John & Srivastava (1999)	5
Meaning in life questionnaire (MLQ)		Steger et al. (2006)	3
Depression anxiety stress scales (DASS-21)		Lovibond & Lovibond (1995)	2
Scales of psychological well-being (PWB)	Various measures of psychological well-being	Ryff (1989)	2
Rosenberg self-esteem scale (RSE)		Rosenberg (1965)	2

2.3.2.2. The AS as an outcome measure and with clinical samples

Out of the 85 samples across 60 publications with AS data, no samples were from clinical populations, nor had any of the publications used it as an outcome measure for psychotherapy.

2.3.3. Results of the meta-analysis

2.3.3.1. Internal reliability (coefficient alpha)

Authenticity Scale. Coefficient alphas for full AS scores were identified for 30 samples ($n = 8823$), with an estimated overall alpha of .84 ($SE = 0.010$) 95% CI [.82, .86]. There was significant heterogeneity between studies reporting alphas for the AS, Cochran's Q test: $\chi^2 (29) = 195.49, p < .001; I^2 = 93.2\%$. Figure 4 shows the forest plot for the AS alphas.

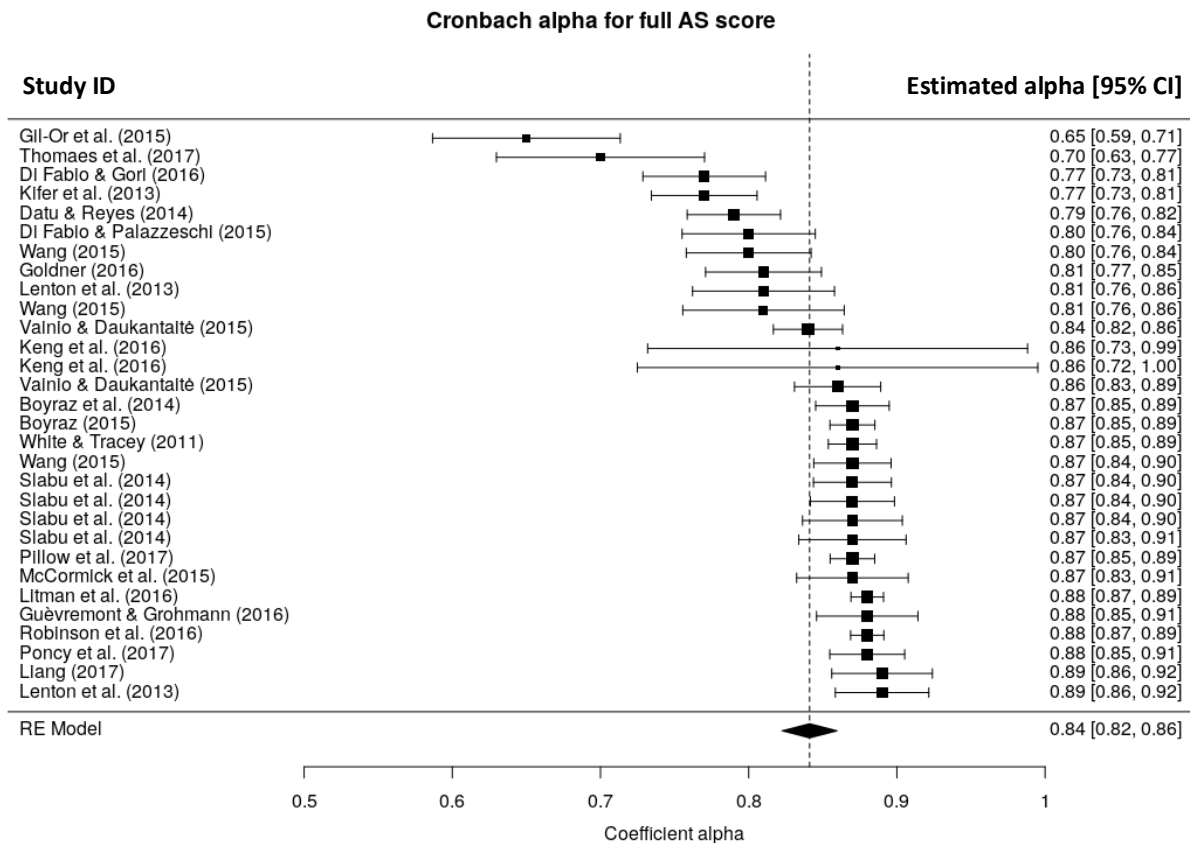


Figure 4. Forest plot of estimated coefficient alphas for the AS.

Figure 5 shows a funnel plot of alphas for the AS against the inverse standard error. The Begg-Mazumdar correlation test of funnel plot asymmetry was significant (Kendall's tau $b = -.56$, $p < .001$).

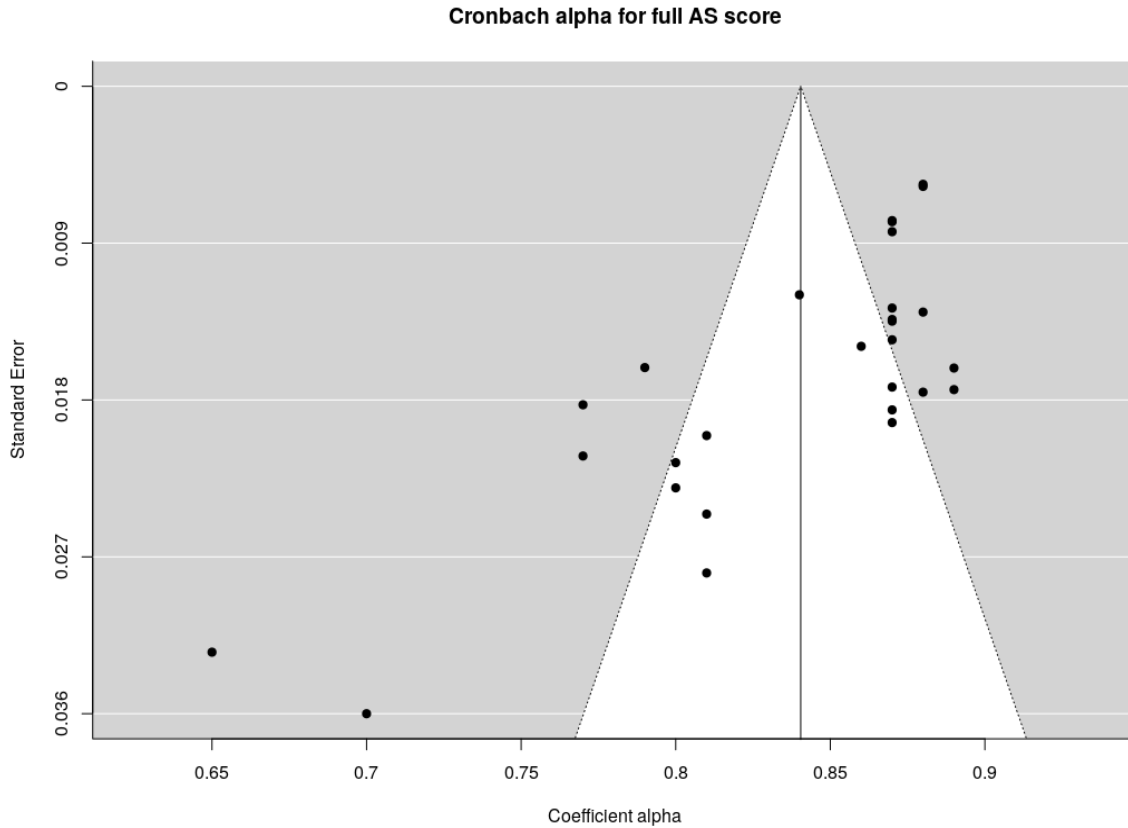


Figure 5. Funnel plot of estimated coefficient alphas for the AS against standard

Subscales. For the AL subscale, alphas from 41 samples were analysed, with an estimated overall alpha of .77 ($SE = 0.011$) 95% CI [.75, .79]. There was a significant heterogeneity between studies reporting alphas for AL, Cochran's Q test: $\chi^2 (40) = 510.16$, $p < .001$; $I^2 = 91.5\%$. The Begg-Mazumdar correlation test of funnel plot asymmetry was significant (Kendall's tau $b = -.29$, $p = .008$).

For the AEI subscale, alphas from 32 samples were analysed, with an estimated overall alpha of .81 ($SE = 0.009$) 95% CI [.80, .83]. There was significant heterogeneity between studies

reporting alphas for AEI, Cochran's Q test: $X^2(31) = 288.92, p < .001; I^2 = 89.2\%$. The Begg-Mazumdar correlation test of funnel plot asymmetry was significant (Kendall's tau $b = -.27, p = .03$).

For the SA subscale, alphas from 33 samples were analysed, with an estimated overall alpha of .82 ($SE = 0.013; 95\% CI = .80, .85$). There was significant heterogeneity between studies reporting alphas for SA, Cochran's Q test: $X^2(32) = 677.38, p < .001; I^2 = 96.2\%$. The Begg-Mazumdar correlation test of funnel plot asymmetry was significant (Kendall's tau $b = -.35, p = .004$). See Appendix N for forest and funnel plots of coefficient alphas for each subscale.

Table 3

Estimated combined Cronbach's alphas for the AS and subscales

Scale	k	Cronbach's alpha	SE	95% CI	
				Lower	Upper
Authenticity Scale	28	.84	0.010	.82	.86
Authentic Living	41	.77	0.011	.75	.79
Accepting External Influences	32	.81	0.009	.80	.83
Self-Alienation	33	.82	0.013	.80	.85

k: number of publications reporting Cronbach's Alpha for each scale.

2.3.3.2. Scale intercorrelations

Authenticity Scale. Correlations between the full AS and its subscales were reported in four samples, with a combined $n = 1,997$ (see figure 6). The overall estimated correlation coefficients showed high correlations between the AS and all its subscales: with the AL, $r(1995) = .64 (p < .001, 95\% CI = .33, .94)$, with the AEI, $r(1995) = -.82 (p < .001, 95\% CI = -.87, -.78)$, and with the SA, $r(1995) = -.81 (p < .001, 95\% CI = -.86, -.77)$. There was significant

heterogeneity between the correlation coefficients for the correlations between the AS and all subscales, Cochran's Q test: $X^2(3) = 23.51 - 333.79, p < .001; I^2 = 87.1 - 99.5\%$.

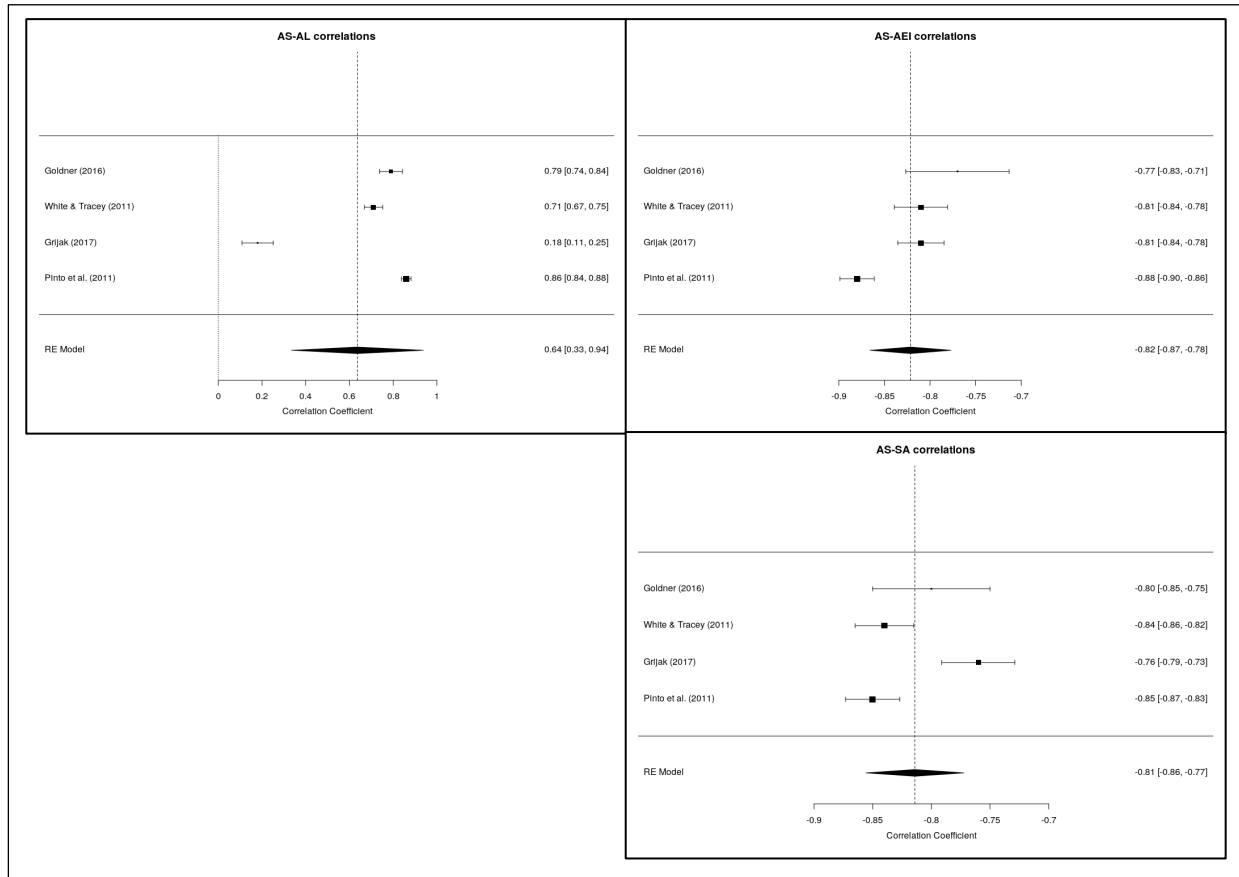


Figure 6. Forest plots of estimated correlation coefficient between the AS and its subscales.

Subscales. Intercorrelations between the AS subscales were reported in 26 samples, with a combined $n = 5,675$ (see figure 7 and table 4). The overall estimated correlation coefficient between the AL and the AEI subscales was $r(5,673) = -.34$ ($p < .001$, 95% CI = $-.39, -.29$), and between the AL and the the SA subscale, $r(5,673) = -.36$ ($p < .001$, 95% CI = $-.41, -.32$). The overall estimated correlation coefficient between the AEI and the SA subscales was $r(5,673) = .44$

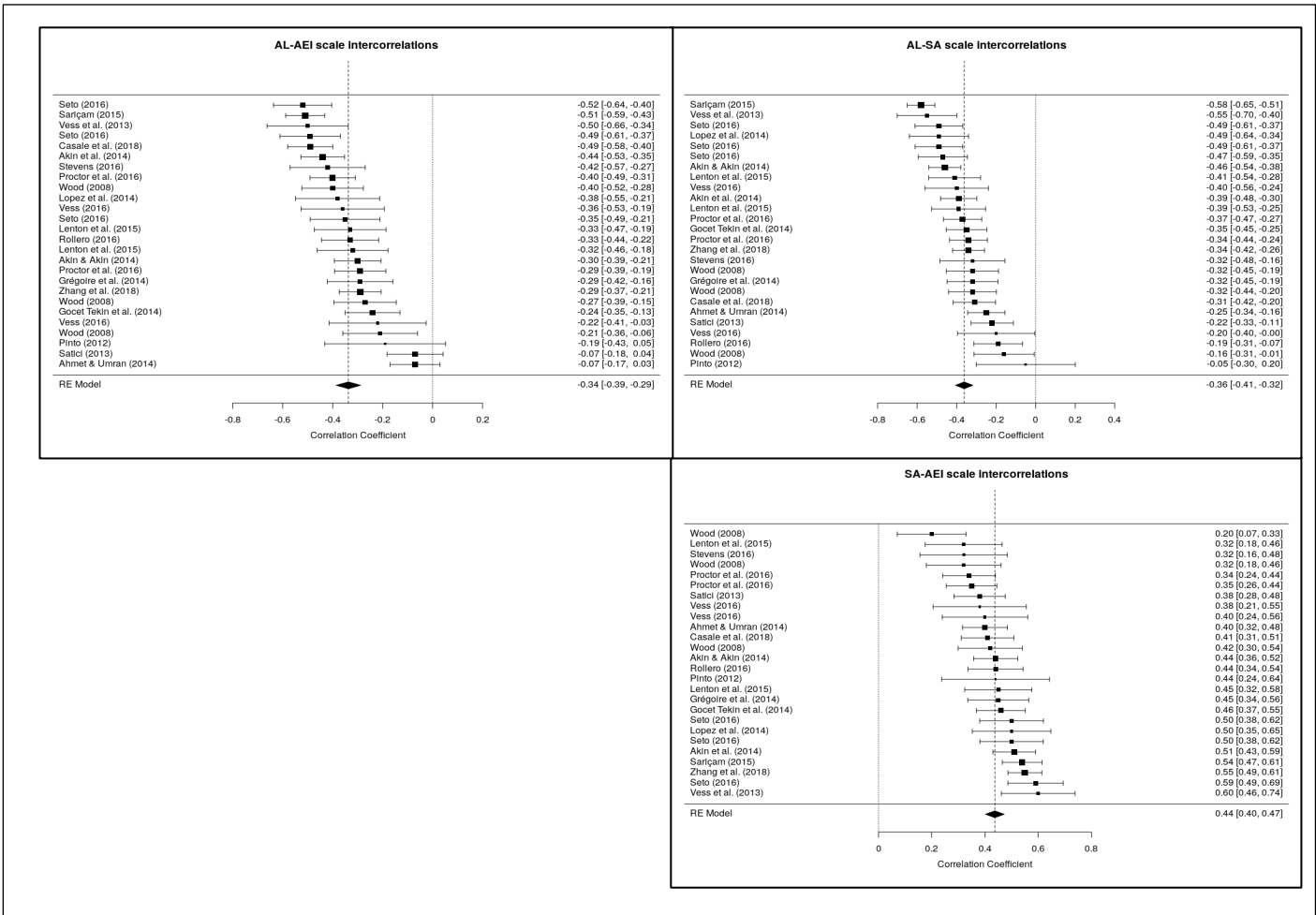


Figure 7. Forest plots of estimated correlation coefficient between the AS subscales.

Table 4

Matrix of the estimated overall intercorrelations for the AS and its subscales

Scale	<i>n</i>	1	2	3	4
1 Authenticity Scale	1997	-	.64	-.82	-.81
2 Authentic Living	5675		-	-.34	-.36
3 Accepting External Influences	5675			-	.44
4 Self-Alienation	5675				-

n = 1997 for correlation coefficients between the AS and its subscales, and *n* = 5675 for subscale intercorrelations

All correlations significant at $p < .001$.

($p < .001$, 95% CI = .40, .47). There was significant heterogeneity between the correlation coefficients for each of the correlations, Cochran's Q test: $X^2(25) = 67.96 - 120.61$, $p < .001$; $I^2 = 63.2 - 77.6\%$.

2.3.3.3. Mean scores

Three studies were excluded from the analysis of means scores because they changed the Likert scoring scale of the AS, from 1-7 to 1-6 or 1-5. A further eight were excluded for reporting erroneous mean data (see section 2.4.2.4).

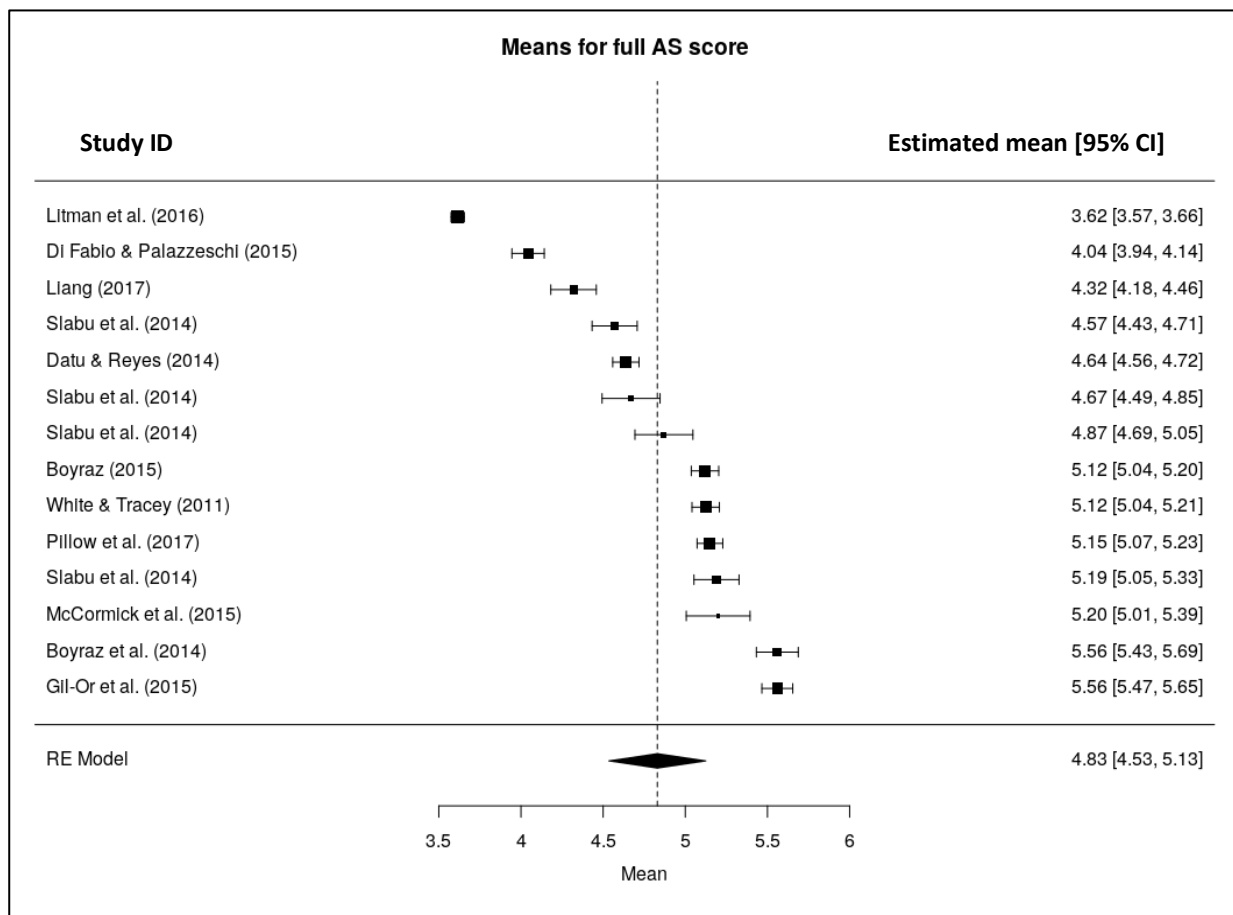


Figure 8. Forest plot of estimated AS means.

Mean Authenticity Scale scores. Mean AS scores from 14 samples were included in the analysis ($n = 4821$). The overall estimated mean was 4.83 ($SE = 0.15$, 95% CI = 4.53, 5.13) (see figure 8), and the POMP score was 63.8%.

There was significant heterogeneity in mean AS scores across studies, Cochran's Q test: $X^2(13) = 3554.51$, $p < .001$; $I^2 = 99.3\%$. Figure 9 shows a funnel plot of estimated means for the AS against the inverse standard error. The Begg-Mazumdar correlation test of funnel plot asymmetry was non-significant, indicating the absence of publication bias (Kendall's tau $b = -.11$, $p = .58$).

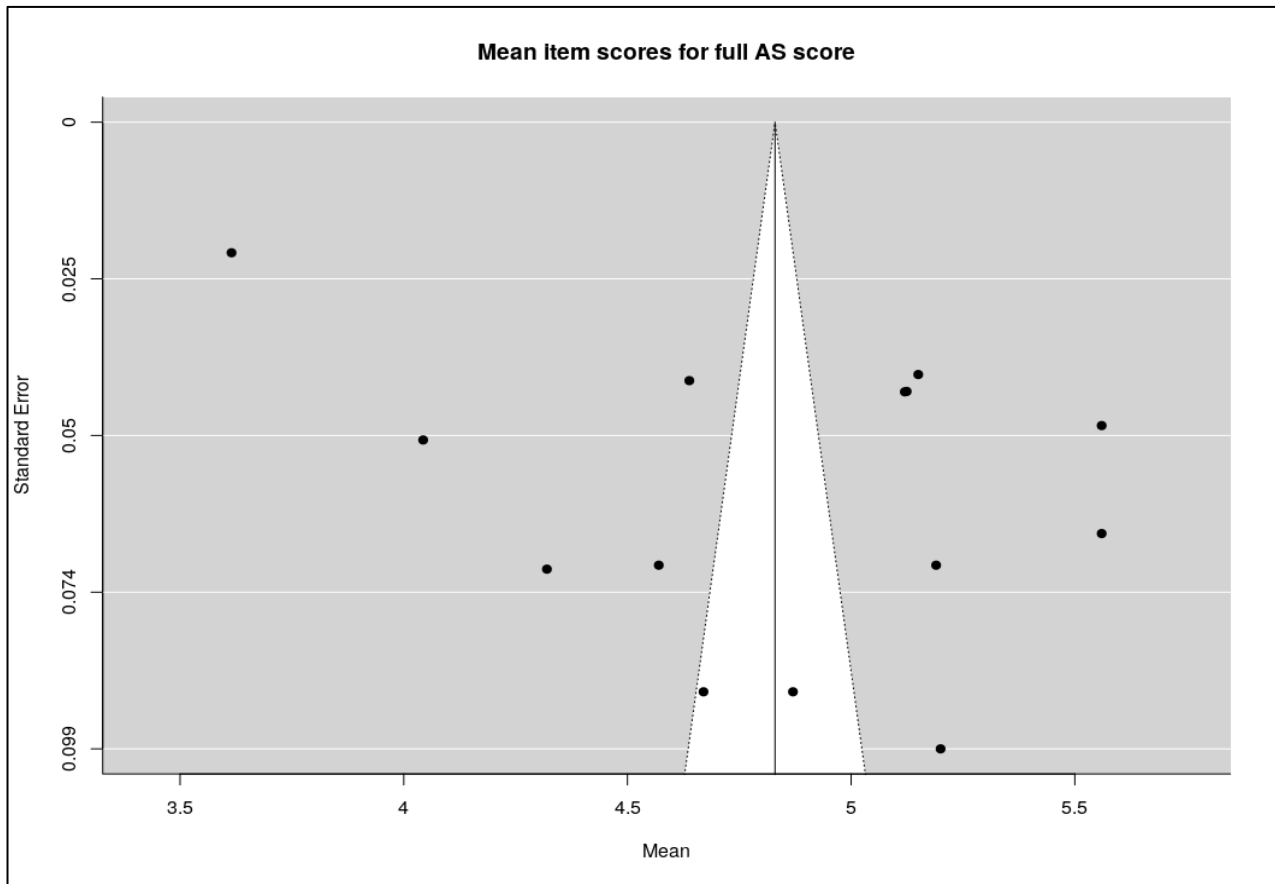


Figure 9. Funnel plot of estimated AS mean scores against standard error.

Mean subscale scores. Mean AL scores from 42 samples were analysed. The overall estimated mean was 5.60 ($SE = 0.05$, 95% CI = 5.50, 5.70). The POMP for the AL was 76.7%. There was significant heterogeneity in mean AL scores across studies, Cochran's Q test: $X^2(41) = 1176.45$, $p < .001$; $I^2 = 96.1\%$.

Mean AEI scores from 38 samples were analysed. The overall estimated mean was 3.47 ($SE = 0.06$, 95% CI = 3.35, 3.59). The POMP for the AEI was 41.2%. There was significant heterogeneity in mean AEI scores across studies, Cochran's Q test: $X^2(37) = 661.19$, $p < .001$; $I^2 = 95.0\%$.

Mean SA scores from 40 samples were analysed. The overall estimated mean was 2.86 ($SE = 0.06$, 95% CI = 2.74, 2.99). The POMP for the SA was 31.0%. There was significant heterogeneity in mean AS scores across studies, Cochran's Q test: $X^2(39) = 893.35$, $p < .001$; $I^2 = 95.3\%$. See Appendix O for forest and funnel plots of mean scores for each subscale.

Table 5

Matrix of the estimated overall means for the AS and its subscales

Scale	k	Mean	95 % CI		SE	POMP
			Lower	Upper		
Authenticity Scale	14	4.83	4.53	5.13	0.15	63.8%
Authentic Living	42	5.60	5.50	5.70	0.05	76.7%
Accepting External Influences	38	3.47	3.35	3.59	0.06	41.2%
Self-Alienation	40	2.86	2.74	2.99	0.06	31.0%

Note. POMP: percent of maximum possible.

2.4. Discussion

2.4.1. Narrative synthesis discussion

2.4.1.1. Concurrent validity

Because of limited space, I grouped the concurrent validity measures by what construct they aim to measure. The two largest groups of measures operationalised aspects of well-being and emotional distress. I address these two first, moving on to the authenticity-related measures, and finally the personality measures. I have included tables for each group giving an overview of the correlations, as there was insufficient space to discuss each one.

2.4.1.2. The AS and measures of well-being

The measure most frequently correlated with the AS in the literature is the Satisfaction with Life Scale (SWLS; Diener *et al.*, 1985), a measure of the evaluative side of subjective well-being, with 5-items developed to measure global cognitive judgement of one's own satisfaction with life. As the AS is based in positive psychology and was itself developed to measure well-being, one would expect significant positive correlations with the SWLS (Wood *et al.*, 2008). Indeed, Wood and colleagues (2008) found the SWLS shared small to moderate positive correlations with the AL subscale in three discrete samples at around $r = .20$, and moderate negative correlations with the SA. Interestingly, the AEI only correlated significantly in one of the three samples ($r = -.35$), suggesting it may be less clearly related to the SWLS. Three other studies (Grégoire, Baron, Ménard, & Lachance, 2014; Proctor *et al.* 2015, Sariçam, 2015) reported similar correlation coefficients, around $r = .25$ for the AL, $r = -.20$ for the AEI, and $r = -.50$ for the SA. For the overall AS, correlation with SWLS was found to be moderate around $r = .35$ in three studies (Boyras *et al.*, 2014; Boyras & Kuhl, 2015; Di Fabio & Palazzeschi, 2015).

Wood *et al.* (2008) also found all three subscales correlated significantly with the RSE (Rosenberg, 1965), a 10-item measure of global self-esteem, with correlations particularly strong for SA (coefficients ranged from $r = -.45$ to $-.59$), whilst correlations were small to moderate for the AL and AEI.

Finally, in three studies the correlations between the positive affect subscales of the PANAS (Watson *et al.*, 1988) and the subscales of the AS were all small to moderate and significant, positive with the AL, and negative with the SA and AEI (Grégoire *et al.*, 2014; Grijak, 2017; Wood *et al.*, 2008). Thus, as would be expected from a well-being measure such as the AS, it seems people who score higher on AL, and lower on SA and AEI tend to score higher on measures of satisfaction with life, self-esteem, and positive affect. Table 6 gives an overview of the correlations between the AS and other well-being measures.

2.4.1.3. The AS predicts scores on the SWLS

Moving beyond cross-sectional designs, Boyraz *et al.* (2014) conducted a longitudinal study to examine whether scores on the AS predicted later life satisfaction, measured by the SWLS. In their study, participants completed the AS and SWLS twice, the second time on average 46.2 days ($SD = 7.39$) after the first. They found that AS scores at baseline significantly predicted life satisfaction at time 2 ($\beta = .37, p < .001$), even after controlling for temporal stability. In contrast, initial scores on the SWLS did not significantly predict AS scores at time 2 ($\beta = -.06, p > .05$). Several studies have already found scores on the AS to be significantly correlated with various measures of well-being (see table 6), but the Boyraz *et al.* (2014) study takes this a step further, finding that the relationship was unidirectional, suggesting that when you become more authentic, your satisfaction with life increases as well. Though, it is important to note that 46 days is not a large prediction interval, this study does appear to be consistent with much of the

psychotherapeutic literature discussed in Chapter 1 and the design model of the AS (e.g. Kohut, 1971; May, 1981; Perls, 1969; Rogers, 1959; Winnicott, 1965).

Table 6
The AS and measures of well-being

Study	Correlates	Correlation coefficients (<i>r</i>)				<i>n</i>	
		Well-being measures	Full AS	AL	AEI		SA
Boyraz & Kuhl (2015)	Life satisfaction		.40**			619	
Grégoire et al. (2014)	Life satisfaction			.23**	-.21**	-.42**	188
Proctor et al. (2016)	Life satisfaction			.27**	-.18**	-.52**	329
Wood et al. (2008)	Life satisfaction			.22**	-.35**	-.34**	180
	Life satisfaction			.22**	-.13 <i>ns</i>	-.50**	158
Robinson et al. (2013)	WEMWBS		.47***				240
Grijak (2017a)	Self-esteem			.32*	-.08 <i>ns</i>	-.18**	206
Wood et al. (2008)	Self-esteem			.23**	-.27**	-.45**	213
	Positive affect			.23**	-.23**	-.21**	180
Grijak (2017a)	Positive affect			.25**	-.13*	-.29**	206
Grégoire et al. (2014)	Positive affect			.21**	-.18*	-.32**	188
Lopez et al. (2015)	Presence of meaning in life			.30***	-.28**	-.44***	158
Vess et al. (2016)	Presence of meaning in life			.27**	-.10 <i>ns</i>	-.52**	93
Wood et al. (2008)	Gratitude			.37**	-.15 <i>ns</i>	-.35**	119
Stevens (2017)	Emotional intelligence			.45**	-.13 <i>ns</i>	-.42**	116
Pillow et al. (2017)	Psychological well-being		.55*				629
Wood et al. (2008)	Self-acceptance			.28**	-.41**	-.39**	180
	Purpose in life			.08 <i>ns</i>	-.22**	-.15*	180
	Positive relationships			.18*	-.27**	-.23**	180
	Personal growth			.25**	-.30**	-.28**	180

Note: **Life satisfaction** = Satisfaction With Life Scale (SWLS; Diener *et al.*, 1985); **WEMWBS** = Warwick-Edinburgh Mental Well-Being Scale (Tennant *et al.*, 2007); **Self-Esteem** = Rosenberg's Self-Esteem Scale (1965); **Positive Affect** = PANAS (Watson *et al.*, 1988); **Presence of meaning in life**: Meaning in life questionnaire (MLQ; Steger *et al.*, 2006); **Gratitude** = The Gratitude Questionnaire-6 (McCullough *et al.*, 2002); **Emotional intelligence** = Self-report emotional intelligence test (SREIT; Salovey & Mayer, 1990); **Psychological well-being subscale, Self-acceptance subscale, Purpose in Life subscale, Positive relations with others subscale, Personal growth subscale** = Ryff's Scales of Psychological Well-being (1989).

* $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$ *ns* = non-significant

2.4.1.4. The AS and measures of emotional distress

Considering the conceptual underpinnings of the AS discussed in chapter 1, it would also be expected to correlate significantly with the opposite end of the well-being spectrum, with measures of emotional distress. According to person-centred theory (Rogers, 1959), this should be particularly true for the SA subscale. Table 7 gives an overview of emotional distress measures correlated with the AS.

Two studies (Boyras & Kuhl, 2015; Boyras, Waits, & Felix, 2014) investigated how the AS relates to the DASS-21 (Lovibond & Lovibond, 1995), a 21-item measure with three subscales measuring depression, anxiety and stress, finding significant large correlations of around $-.50$ with the full AS.

Wood *et al.* (2008) reported correlations between the AS subscales and two measures of anxiety (the Tension subscale of the Profile of Mood States; McNair, Lorr, & Droppleman, 1971) and stress (The Perceived Stress Scale; Cohen & Williamson, 1988). SA shared moderate to large positive correlations with both measures, the AEI shared moderate positive correlations with both measures, whilst the AL only correlated with stress at about $r = -.20$.

Stevens (2017) investigated how the Toronto alexithymia scale (TAS; Bagby, Parker, & Taylor, 1994) relates to the AS. He described alexithymia as ‘a deficiency in understanding, processing and describing emotions’ (Stevens, 2017, p.393) and found it correlated with SA at $r = .61$, but not with either the AL or AEI. This is unsurprising and further supports the construct validity of the AS, as the SA aims to measure the extent of disconnection between awareness and what is happening at a deep level inside, whereas the AL and AEI do not directly relate to affective functioning (Wood *et al.*, 2008).

Thus, the AS is clearly related to measures of emotional distress, as expected, and it is the SA that generally seems most strongly related to most of the measures, which fits with its conceptual underpinnings (Roger, 1959).

Table 7
The AS and measures of emotional distress

Study	Correlates	Correlation coefficients (<i>r</i>)				<i>n</i>	
		Emotional distress measures	Full AS	AL	AEI		SA
Boyraz & Kuhl (2015)	Depression, anxiety, stress		-.55**			619	
Boyraz et al. (2014)	Depression, anxiety, stress		-.49**			232	
Grégoire et al. (2014)	Negative affect			-.15*	.33**	.41**	188
Grijak (2017)	Negative affect			-.10 <i>ns</i>	.03 <i>ns</i>	.25**	206
Wood et al. (2008)	Negative affect			-.07 <i>ns</i>	.20**	.21**	180
	Stress			-.20**	.26**	.47**	158
	Anxiety			-.18 <i>ns</i>	.20*	.39**	104
Pinto et al. (2012)	Aggressive behaviour			-.33*	-.11 <i>ns</i>	-.01 <i>ns</i>	62
	Trait anger			-.04 <i>ns</i>	.10 <i>ns</i>	.30*	62
Stevens (2017)	Alexithymia			-.18 <i>ns</i>	.15 <i>ns</i>	-.61**	116
	Avoidant attachment style			-.07 <i>ns</i>	-.01 <i>ns</i>	.41**	116
Goldner (2016)	Avoidant attachment style	-.27***		-.33***	.13 <i>ns</i>	.19**	200
Lopez et al. (2015)	Avoidant attachment style			-.26*	.23*	.24*	100
	Anxious attachment style			-.34***	.47***	.57***	100
Stevens (2017)	Anxious attachment style			-.06 <i>ns</i>	-.29**	-.29**	116
Goldner (2016)	Anxious attachment style	-.33***		-.23**	.34***	.21**	200

Note: **Aggressive behaviour** = Point subtraction aggressive paradigm (PSAP; Cherek, 1981); **Alexithymia**: Toronto alexithymia scale (TAS; Bagby *et al.*, 1994); **Anxiety** = Self-Rating Anxiety Scale (SAS; Zung, 1971); **Anxious and avoidant attachment style**: Experiences in close relationships (ECR; Brennan *et al.*, 1998); **Depression, Anxiety, Stress** = Depression, Anxiety, Stress Scales-21 (DASS-21; Lovibond & Lovibond, 1995); **Negative Affect** = PANAS (Watson *et al.*, 1988); **Stress** = The Perceived Stress Scale (Cohen & Williamson, 1988); **Trait anger**: anger facet IPIP NEO-PI R (Goldberg *et al.*, 2006).

p* ≤ .05 *p* ≤ .01 ****p* ≤ .001 *ns* = non-significant

2.4.1.5. The AS and other measures of authenticity

An important part of exploring the concurrent validity of the AS is looking at how it relates to other measures of authenticity. One such measure is the AI-3 (Kernis & Goldman, 2006; see section 1.3.9.4), which is one of the measures most closely related to the AS, as it also aims to

measures trait authenticity. Unsurprisingly, all four subscales of the AI-3 shared moderate to large correlations with all three subscales of the AS in expected directions, apart from two that were non-significant: AL to Unbiased processing and AEI to Relational orientation. The second of these makes conceptual sense: being relationally orientated would not necessary be strongly related to how much an individual is influenced by others. It is however, surprising that Unbiased processing did not correlate with the AL. The subscales that seem to overlap the most conceptually are Awareness and the SA, and Behavioural and the AL, and indeed, both these correlations were large.

I also included two other measures in this group, the organismic valuing scale (Govindji & Linley, 2007) and the self-concept clarity scale (SCCS; Campbell, Trapnell, Heine, Katz, Lavalley, & Lehman, 1996), both of which are very closely linked to the person-centred definition of authenticity (Barrett-Lennard, 1998). As expected, both of these shared moderate to large correlations with the AS subscales (see table 8).

Table 8
The AS and measures of authenticity

Study	Correlates Measures of authenticity	Correlation coefficients (r)				n
		Full AS	AL	AEI	SA	
Stevens (2017)	AI-3: Awareness	.39**		-.31**	-.62**	116
	AI-3: Unbiased processing	.06 <i>ns</i>		-.32**	-.39**	116
	AI-3: Behavioural	.65**		-.51**	-.42**	116
	AI-3: Relational orientation	.40**		-.15 <i>ns</i>	-.45**	116
Proctor et al. (2015)	Organismic valuing	.51**		-.28**	-.64**	329
Vess et al. (2016)	Self-concept clarity	.22*		-.30**	-.68**	93

Note: **AI-3 subscales** = Authenticity Inventory-3 (Kernis & Goldman, 2006); **Organismic Valuing**: Organismic valuing scale (Govindji & Linley, 2007); **Self-concept clarity** = Self-concept clarity scale (SCCS; Campbell et al., 1996).

* $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$ *ns* = non-significant

2.4.1.6. The AS and measures of personality

In Chapter 1 I addressed the debate in the literature regarding the state and trait views of authenticity (e.g. Lenton, Bruder, *et al.*, 2013). The AS is firmly grounded in the trait view, seeing it as a relatively stable, base-rate propensity towards behaviour and cognition that is in accordance with a ‘true self’, though by no means unchangeable (Wood *et al.*, 2008). Because of this, Wood and colleagues (2008) were interested in how it relates to other stable traits, such as measures of personality. They investigated how the AS relates to the Big Five Inventory (BFI; John & Srivastava, 1999), a measure of personality traits using the five-factor model of personality (McCrae & John, 1992). They were particularly interested in what proportion of the variance in AS scores could be explained by BFI scores, as high amounts of variance explained by the BFI might imply the AS is simply measuring a combination of other personality traits (Wood *et al.*, 2008). They expected measures of personality to only relate moderately to the AS, thus supporting its concurrent validity. Consistent with their hypothesis, regression analysis showed that the BFI accounted for a small, but significant portion of the variance in AS scores, between 11% and 13% for each of the AS subscales. In developing the French translation of the AS, Grégoire *et al.* (2014) found that Big-five traits significantly explained for 26% of the variance in both AEI and SA scores, but none of the variance in AL scores. Both studies concluded that the AS is meaningfully related to, but distinct from, the Big-five traits, and that more authentic people tend to be more extraverted, agreeable, conscientious, open and less neurotic (see table 9).

Pinto, Maltby and Wood (2011) further investigated the concurrent validity of the AS by exploring its relationship with Gray’s (1982) behavioural inhibition/activation systems, measured by the behavioral inhibition and activation scales (BIS/BAS; Carver & White, 1994), and Cloninger’s psychobiological model (Cloninger, Svrakic, & Przybeck, 1993). The latter was

developed to help predict vulnerability to mental disorders, particularly personality disorders (Cloninger, *et al.*, 1993) and was measured by items on the international personality item pool (TCI-IPIP, Goldberg *et al.*, 2006).

The BIS subscale showed a small negative correlation with the full AS ($r = -.24, p < .001$), and similarly correlations with all three of its subscales, whereas BAS did not correlate significantly with any of these. Regression analysis showed that only the BIS subscale ‘Anxiety’ significantly predicted AS scores, accounting for 5.7% of the variance (Pinto *et al.*, 2011). Of the dimensions in Cloninger’s biopsychological model, Pinto *et al.* (2011) found that harm-avoidance, reward-dependence, self-directedness, persistence and self-transience all weakly predicted AS scores, but together only accounted for 18% of the variance.

Thus, it seems that though measures of all three personality models discussed above were related to the AS, except for the BAS, none of the measures accounted for a large degree of the variance in AS scores, indicating the AS does not simply measure a combination of personality traits (Wood *et al.*, 2008).

Table 9

The AS and measures of personality

Study	Correlate	Correlation coefficients (<i>r</i>)				<i>n</i>
		Full AS	AL	AEI	SA	
	<i>Big Five</i>					
Wood et al. (2008)	Extraversion	.11 <i>ns</i>	-.33**	-.29**	94	
	Agreeableness	.27**	-.07 <i>ns</i>	-.24*	94	
	Conscientiousness	.17 <i>ns</i>	-.16 <i>ns</i>	-.28**	94	
	Neuroticism	-.23*	.19 <i>ns</i>	.25*	94	
	Openness to experience	.21*	.10 <i>ns</i>	-.04 <i>ns</i>	94	
Grégoire et al. (2014)	Extraversion	.05 <i>ns</i>	-.18*	-.36**	188	
	Agreeableness	.27**	-.07 <i>ns</i>	-.24*	188	
	Conscientiousness	.34**	-.34**	-.29**	188	

	Neuroticism		-.13 <i>ns</i>	.50**	.50**	188
	Openness to experience		.06 <i>ns</i>	.13 <i>ns</i>	-.26**	188
	<i>BIS/BAS</i>	<i>Full AS</i>	<i>AL</i>	<i>AEI</i>	<i>SA</i>	
Pinto et al. (2011)	Behavioural inhibition (BIS)	-.24***	-.20***	.23***	.20***	554
	Behavioural activation (BAS)	.04 <i>ns</i>	-.01 <i>ns</i>	-.08 <i>ns</i>	-.03 <i>ns</i>	554
	<i>TCI-IPIP</i>	<i>Full AS</i>	<i>AL</i>	<i>AEI</i>	<i>SA</i>	
Pinto et al. (2011)	Novelty-seeking	.05 <i>ns</i>	.03 <i>ns</i>	-.05 <i>ns</i>	-.05 <i>ns</i>	554
	Harm-avoidance	-.22***	-.17***	.19***	.21***	554
	Reward-dependence	.26***	.29***	-.17***	-.21***	554
	Persistence	.11***	.12***	-.08 <i>ns</i>	-.09*	554
	Self-directedness	.17***	.15***	-.14***	-.15***	554
	Cooperativeness	-.02 <i>ns</i>	-.03 <i>ns</i>	.00 <i>ns</i>	.04 <i>ns</i>	554
	Self-transcendence	.21***	.23***	-.16***	-.15***	554

Note: Big Five: Big Five Inventory (John & Srivastava, 1999); **BIS/BAS:** Behavioural Inhibition/Activation scales (Carver & White, 1994); **TCI-IPIP:** Cloninger's temperament and character items from the international personality item pool (Goldberg *et al.*, 2006).

* $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$ *ns* = non-significant

2.4.1.7. The AS with clinical samples

Despite the systematic literature search identifying 60 studies that reported quantitative data on the AS, it failed to find any studies that have used the AS as an outcome measure for psychotherapy to date. Furthermore, I found that none of the identified publications have used it with clinical samples.

2.4.2. Meta-analysis discussion

The meta-analysis aimed to summarise the psychometric properties of the AS and its subscales, collating data from all relevant studies. Data from 60 publications was synthesised, with a combined sample size of 23,043.

2.4.2.1. Reliability generalisation

Vacha-Haase (1998) advocated meta-analysing reliability values, terming it reliability generalisation. This involves exploring the properties of coefficient alphas for a specific measure across studies, looking at typical coefficient alpha values, and pivotally the variability of these values, along with the sources of this variability (Vacha-Haase, 1998). As reliability is a property of the scores of a particular sample, rather than a property of the measure itself, reliability generalisation uses a random-effects model, as discussed earlier, and the exploration of variance in coefficient alphas is a pivotal element along with calculating an overall estimate (Bonett, 2010; Bornstein *et al.*, 2010).

In this meta-analysis, coefficient alphas from all available studies were synthesised in a reliability generalisation, calculating overall estimated coefficient alphas for both the AS and its subscales, with 95% confidence intervals (CI). Although, results from the majority of papers were analysable, it is important to note that 7 out of the 60 (11.7%) studies that reported scores for the AS or its subscales did not report coefficient alpha for their samples, but instead reported internal reliability from a different study (with a different sample), or made no reference to reliability at all. As has been repeated numerous times in the psychometric literature (e.g. DeVellis, 2012, Vacha-Haase, 1998), there is no reason not to report internal reliability of any multi-item scale for the reported sample, and citing other reliability values is no substitute for this.

Variance in the combined coefficient alphas. As expected, the between-study variance in reported coefficient alphas for AS scores was very high, with Cochran's Q test significant and the I^2 statistic showing 93.2% variance. For the subscales, the I^2 statistic ranged from 88.7% to 96.1%. Though this degree of heterogeneity is unsurprising, considering the vastly different samples included in the meta-analysis, it is important and interesting, essentially indicating that the overall

coefficient alpha estimates of $\alpha = .84$ (95% CI [.82, .86]) for the AS cannot be generalised beyond this study. Similar degrees of heterogeneity were found for the combined coefficient alphas of the three subscales, which were estimated at between $\alpha = .77$ and $\alpha = .82$. In an attempt to make sense of this considerable heterogeneity, I tried categorising samples, but no clear pattern emerged that might account for the variance.

The heterogeneity in AS coefficient alphas can be explored by inspecting the forest plot (see figure 4). At the lower end, five of the studies had upper 95% confidence limits (CL) that fell below the estimated aggregate of $\alpha = .84$. The plot suggests the two lowest of these were outliers, with coefficient alphas considerably lower than the next lowest value. Indeed, coefficient alphas from the third lowest to the highest ranged by .12, from $\alpha = .77$ to .89, whereas the range from the third lowest to the lowest alone was .12, from $\alpha = .77$ to .65. At the other end of the plot, 10 studies had lower 95% CLs above the overall aggregate. In general, the studies with higher coefficient alphas had narrower 95% CIs than the studies with lower values.

Taken together, there is nothing in this reliability generalisation indicating the internal reliability of AS scores in other populations will not be good enough to make the measure useable. However, the high levels of variability found here serves to underline the unquestionable importance of estimating and reporting internal reliability for every sample.

Publication bias. The Begg-mazumdar correlation test was significant for the coefficient alphas for AS scores, indicating the potential presence of publication bias. However, the correlation was negative, which indicates the asymmetry was to the left of the funnel plot. This is clear when inspecting the plot, showing the asymmetry does not stem from too many high coefficient alphas published, but rather two alphas considerably lower than the mean. In this case then, the asymmetry of the funnel plot does not seem to indicate publication bias.

2.4.2.2. Scale intercorrelations

The estimated overall correlation coefficients between the subscales and the overall scale were all high, ranging from .64 to .82. However, it is important to note that assuming approximately equal variance across items, 33% of the variance of the total score is shared with each subscale, which inflates these correlations markedly. The directions of these correlations were all as expected, with the full AS scale and the AL subscale positively worded (high scores indicating high authenticity) and the AEI and SA subscales negatively worded.

The estimated intercorrelation coefficients between the subscales were all moderate, ranging from .34 to .44. Though the directions of the reported correlations were all as expected, three of the AL to AEI coefficients had 95% CIs that crossed 0, and so did two of the AL to SA coefficients. Taken together with the high combined coefficient alpha values of the subscales, these findings provisionally support the structure of the design model reported by Wood *et al.* (2008), namely that the three subscales were developed to measure three discrete constructs that together form the humanistic definition of authenticity (Barrett-Lennard, 1998; Rogers, 1959).

One correlation coefficient between the AS and the AL subscale was a clear outlier, reported at $r = .18$ (Grijak, 2017a). This correlation was observed in the Serbian translation of the scale rather than the English version, but translation seems unlikely to be the cause of the discrepancy, particularly as the other two subscales correlated to the expected degrees with the full scale and with AL. This correlation coefficient is so at odds with coefficients from the other samples reviewed that it seems likely to be artefactual, particularly considering the reported mean scores for all the subscales and the full scale in the study were similar to the estimated overall mean scores in this meta-analysis. The author of the study has been contacted for clarification, but has not responded.

2.4.2.3. Mean scores and distributions

The estimated overall mean for the AS was 4.83, which is 63.4% of the maximum score, 7, showing that the estimated mean AS score was close to the middle of the scoring scale. This is ideal, suggesting the absence of any floor or ceiling effects for AS scores, though *SDs* would also be needed to confirm this. If floor and ceiling effects are present, which would likely manifest as POMP scores close to 90% or 10%, it indicates that only a part of the scoring scale has been used, which in turn means there is relatively little variance in scores. As correlation is a function of variance, the presence of floor or ceiling effects would reduce the power of any correlational tests, limiting the usefulness of a measure (Cramer & Howitt, 2004).

Considering the three subscales separately, their POMP scores diverged considerably from the overall scale. Most markedly, the AL subscale had a mean score of 5.60, which is a POMP of 76.7%. The AEI had an estimated mean score of 3.47 and a POMP of 41.2%, which is fairly close to the middle of the scoring scale, whilst the SA estimated mean score of 2.86 is a POMP of 31.0%. This shows that respondents tend to rate themselves considerably higher on the AL than on the AEI and SA, particularly the SA. One possible reason for this is social desirability. The AL is the only positively worded subscale, with items such as: “I live in accordance with my values and believes” (item 12, Wood *et al.*, 2008), and it is reasonable to expect that respondents would wish to score higher on such an item than on the SA subscale with items such as: “I feel as if I don’t know myself very well.” (item 7, Wood *et al.*, 2008). The AEI subscale seems to be worded more neutrally than the AL and SA, with items such as: “Other people influence me greatly.” (item 6, Wood *et al.*, 2008), which might account for why its scores were fairly close to the middle of the scoring scale.

However, Wood and colleagues (2008) did investigate whether the AS was affected by social desirability when developing it, by including the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1984) as a discriminant validity measure. They found the correlations to be very small and non-significant for both the overall scale and each of the subscales, and thus concluded the AS is not meaningfully affected by socially desirable responding. It is worth noting that the mean scores of the sample ($n = 180$) with which Wood *et al.* (2008) performed the correlation with the BDRI differed from the estimated means in this meta-analysis, particularly the scores on the AL subscale, which had a POMP of 64.7% in the Wood *et al.* (2008) sample, 12% lower than the POMP in this meta-analysis. Thus it is possible the AL subscale would have been found to be significantly affected by socially desirable responding if examined in a different sample.

Another important point that needs considering, is the effect of including both negatively and positively worded subscales in the AS. Whereas this might help mitigate against acquiescent response bias, Roszkowski and Soven (2010) suggest it may introduce artefacts, with some research showing that the direction of the wording of items and scales affects score distributions (e.g. Barnette, 1996; Van Dam, Hobkirk, Danoff-Burg, & Earleywine, 2012), and that including both positively and negatively worded subscales may lower, rather than raise, the measure's validity. It may be argued then that either wording all AS items in the same direction, or including both positive and negative items in each subscale, would make the AS more psychometrically sound.

Variance. As with the coefficient alphas, there was a very high degree of variance in mean AS scores across samples ($I^2 = 99.3\%$). Indeed, the means in the separate samples ranged from 3.62 (POMP of 43.7%) to 5.56 (POMP of 76.0%). As discussed earlier, this level of heterogeneity

is not surprising, considering I am combining mean scores from a psychometric measure expected to vary between individuals and samples. Regardless, I tried categorising samples to find the source of this variability, but no clear pattern emerged. Most samples were from the US or the UK, but their scores did not differ systematically from samples from other countries. No samples could be logically expected to have lower AS scores, and though one sample was of psychotherapists whom one might expect to have high scores, their mean scores were close to the subscale means.

Part of the variance is likely due to the confusion regarding how to report the negatively worded subscales AEI and SA. In their original paper, Wood *et al.* (2008) reported the means of the SA and AEI prior to recoding them, and thus high scores on these two subscales indicated low authenticity. However, as none of the papers in this meta-analysis clarify whether they are reporting recoded means or not, it is only possible to tell what is reported in the very few publications that report both the total AS score and the subscale scores. It is likely that the majority have done as advocated by Wood *et al.* (2008) in the original paper, but I identified three papers that reported the recoded means for the AEI and SA, and there are likely more. Regardless, the large degree of heterogeneity between samples means that the overall estimated means cannot be generalised to any other samples.

Publication bias. In contrast to the funnel plot for AS score coefficient alphas, the Begg-Mazumdar test for plot asymmetry was non-significant ($p = .26$), indicating the absence of publication bias. However, it should be noted that publications adding together raw scores, or reporting recoded AEI and SA scores as if they were recoded, will mitigate against the test detecting a publications bias. Regardless, considering the samples reporting means ($k = 15$) and coefficient alphas ($k = 29$) are largely overlapping, and neither coefficient alphas or means have a clear cutoff point that might dissuade researchers from publishing, such as the significant/non-

significant p -value, it seems reasonable to conclude that there was no publication bias concerning AS means, coefficient alphas, or correlation coefficients.

2.4.2.4. Limitations to the meta-analysis

A significant limitation of the meta-analysis is the uncertainty around the estimated overall means, due to the confusion in the literature around whether to report raw or recoded scores for the AEI and SA subscales, and similarly, whether to simply add together the raw subscale scores, or to recode them first. I identified three papers that had certainly reported recoded means for AEI and SA, two more where the subscale scores did not add up to the reported full AS score, and a further three that certainly added the subscales together without recoding AEI and SA, with two more I suspect have done the same. Because of these examples of poorly or wrongly reported data, I was particularly vigilant reading the papers included in the meta-analysis. In most cases, the authors had either reported both full scale and subscale scores, or subscale intercorrelations, allowing me to see if the means were recoded or not. Of the publications that had not reported either of those things, most had reported coefficient alpha, and as simply adding together the raw scores would greatly decrease the alpha value (one study that did this reported $\alpha = .53$), I felt reasonably confident these studies had recoded the AEI and SA correctly before adding the together.

The ideal way to deal with this issue would have been to contact every author to clarify what had been done, but that fell beyond the scope of this thesis. Instead, I have removed the studies that have clearly reported erroneous data, corrected the data where possible, and assumed that the remaining publications where there is no way to tell what has been done, have correctly added together the subscales.

A second point limitation is that I was not able to trace the considerable between-studies variance to any variables, beyond my attempt at categorising the samples. It would have been very interesting to test whether age, gender, or country accounted for any of the heterogeneity.

2.5. Conclusion and implications for Study 2

In this systematic review, I aimed to summarise the psychometric properties of the AS from all publications that have reported AS data. I addressed these aims in two general parts: a narrative synthesis and a meta-analysis. The narrative synthesis explored the concurrent validity of the AS, reviewing correlations between the AS and measures of well-being, emotional distress, authenticity, and personality. I found that the vast majority of correlations were consistent with the design model of the AS, and its conceptual underpinnings, supporting its validity. The narrative synthesis also revealed that no studies to date have used the AS as an outcome measure, and that no studies have used it with clinical samples.

The meta-analysis used a random-effects model to combine AS coefficient alphas, mean scores and subscale intercorrelation coefficients. It found that for the vast majority of reported samples, the AS showed high internal reliability, intercorrelations consistent with the design model of the AS, and mean scores close to the middle of the scoring scale, qualities needed in an outcome measure. There was however, a very large degree of between-study heterogeneity. Though this was expected considering the effects that were being combined, it is an important point to note when generalising from these findings.

In all, the systematic review found considerable support for the construct validity of the AS, and that it had excellent psychometric properties in the vast majority of relevant samples. However, as none of the identified studies had used the AS with clinical samples, how the AS

would actually function as an outcome measure remains a gap in the literature, one the second study in this thesis is the first to address.

STUDY 2

Chapter 3. Methods

3.1. Introduction

In this chapter, I aim to provide an overview of how I have conducted Study 2. I start by discussing my epistemological stance, a natural starting point for any research. I then review the aims of this thesis, and how Study 2 will address these, before moving on to describe the different elements of the research, starting with its design, a description of the participating counselling services, therapists and clients, and the research procedure. Finally, I go through each of the a priori planned analyses that I will use to answer the research questions, introducing each and briefly discussing the implications of the statistical concepts and methods.

3.2. Epistemology

The perhaps most fundamental questions when conducting research, are those of ontology and epistemology. Ontology is the theory of reality. What is the nature of reality, what exists? Closely linked is epistemology, the theory of knowledge. How do we know what we know? All research and research methodology makes explicit and implicit assumptions regarding these questions (Blaikie, 2004). As a researcher, I believe it is important to explicate my position, thus making transparent the assumptions that underlie this research.

In this thesis, I draw on several different epistemological traditions. I will first discuss empiricism, positivism, and realism, which are closely related, and which lay the foundation for measurement in psychology as a practice. I then briefly discuss constructivist positions, before

moving on to review pragmatism. Finally, I discuss my primary position for this research, pragmatism-realism, and why I see it as the most appropriate for measurement in psychology.

3.2.1. Empiricism and positivism

Empiricism was represented by philosophers such as Locke and Hume in the 18th century, seeing direct observation and sensory experience as the basis of all knowledge (Maul, Wilson, & Irribarra, 2013). Closely related is positivism, a term coined by Comte (1974) in the 19th century, who held that the current age should be one where we apply knowledge from the natural sciences positively to drive progress, both technological and medical, but also societal (Cruickshank, 2012). Logical positivism, which emerged in the 1920s, builds on, and is largely overlapping with, empiricism, but combined it with recent advances in mathematics and logic (Hammersley, 2011). The central doctrine of this epistemology is that all meaningful knowledge is either empirical or logical. Thus, knowledge comes solely from employing strict scientific methodology to our observations of the world, that is by measuring it, and by using logical reasoning to extrapolate from these measurements (Hammersley, 2011).

Measurement in psychology has historically based itself in empiricism and positivism (Guyon, Kop, Juhel, & Falissard, 2018). The issue however, is that the constructs we seek to measure in psychology cannot be directly observed. Thus, in order to measure a latent construct, it must first be operationalised by a psychometric measure. For example, depression as a construct cannot be empirically observed unless it has first been operationalised, as with the BDI (Beck *et al.*, 1961). We might say someone is depressed, but until we list a series of markers and score people on those, we cannot say who is depressed and who is not, or how depressed a person is. The problem is that, embedded in psychometric measurement scores, is random and systematic error (Guyon *et al.*, 2018). Thus a major limitation in holding to logical positivism in psychological

measurement is that an unknown proportion of any score will be error, which means we cannot know whether we have successfully operationalised the construct (Sijtsma, 2012).

3.2.2. Realism

Ontologically, realism shares the view of empiricism and positivism, in that it sees reality as objective and independent from the observer experiencing it, but accurately symbolised in our perception of it (Boyd, 1983; Payne & Payne, 2004). However, where empiricism and positivism hold that observable phenomena make up the totality of reality, realism argues unobservable causal laws interact to bring about the change that can be observed (Bhaskar, 1997; Boyd, 1990; Payne & Payne, 2004). Realism is thus a post-positivist epistemology, aiming to describe the underlying phenomena that predict the observable world (Cruickshank, 2012; Payne & Payne, 2004). Moreover, because elements of the real world cannot be observed directly, realism sees scientific theories as fallible and only an approximation of reality, and in need of constant revising to better describe it (Bhaskar, 1997; Boyd, 1983; Long, 2013; Steinmetz, 1998).

At a glance, realism seems a more suited epistemology for psychological measurement and this thesis, than positivism. It accounts for the proximal nature of a measure, such as the AS, in operationalising authenticity, recognising the inevitability of error being embedded in psychometrics. However, realism holds that reality is 'logically independent of any knower' (Michell, 2003, p. 17). In other words, reality is mind-independent, and does not depend on our awareness of it (Maul *et al.*, 2013). This makes sense when considering natural phenomena: leaves rustling in a forest make sound, regardless of whether or not someone is there to hear it. However, the implication for psychological measurement is the constructs must be ontologically objective, independent of our awareness of them, in order from them to be measurable. In this, realism does not seem compatible with the concept of psychological constructs. Authenticity is not simply an

objective, yet unobservable, phenomenon, causing ripples in the observable world. Rather, authenticity, and any psychological construct, depends on our definition of it. Thus, a purely realist epistemology does not seem reconcilable with the idea of psychological measurement (Maul *et al.*, 2013).

3.2.3. Constructivism

Occupying the opposite end of the epistemology spectrum to positivist and realist paradigms, are constructivist positions. These cover an array of epistemologies, but a common-denominator is the belief that the world cannot be known directly or objectively. Instead, each person mentally constructs the world of experience, and thus knowledge, through their individual cognitive processes (Young & Collin, 2004). In other words, whereas realism holds we have knowledge of reality, constructivism argues we only have knowledge of a construction of reality, a proxy. Within constructivism, there are three general positions. Radical constructivist, such as von Glaserfield (1995), hold that there is no external reality, but that reality is constructed in the mind of the individual. Moderate constructivists on the other hand, such as Piaget (1969), see the individual's constructions of reality as being built in a systematic relationship with the external world. Finally, social constructivists, such as Vygotsky (1978), emphasise the intersubjective, believing the individual cognitive processes that construct reality are preceded by, and derived from, social relationships (Young & Collin, 2004).

From a constructivist epistemology, psychological measurement, and collating measurements from a group of people, appears very problematic. Central to the notion of defining authenticity, reducing it to a set of markers, measuring how people score on those markers, and then combining those measurements, is that there is an objective reality, and thus that one person's AS scores can be meaningfully combined with those of another person. This however, is

fundamentally incompatible with the belief that the world is cognitively constructed by each individual.

3.2.4. Pragmatism

Pragmatism is based on an ontological agnosticism, claiming that the scientific methods we use may not be as objective and independent as the positivists or realists might claim (Long, 2013). As an example, for several centuries scientific development was required to be coherent with the Catholic Church, resulting in claims that today seem outlandish (Kuhn, 1962). Pragmatism is not necessarily at odds with realism, as long as the realist position in question is not too extreme, and holds there may be an objective reality, but our knowledge of reality is shaped and organised by social context (Guyon *et al.*, 2018). Thus, it may be seen as occupying space somewhere between realism and social constructivism. Pragmatism covers an array of positions, but a common denominator is an orientation towards practice, aiming to interpret each notion according to its practical consequences (Maul *et al.*, 2013). James (1907) stated: ‘What difference would it practically make to anyone if this notion rather than that notion were true? If no practical difference whatever can be traced, then the alternatives mean practically the same thing, and all dispute is idle’ (p.14). Pragmatism sees our beliefs and theories about the world as tied to our practical and pragmatic engagements in it, rather than how the world actually is (Guyon *et al.* 2018). When considering measuring authenticity, a pragmatist might be less concerned with the ontology of authenticity, or how the AS operationalises it, but rather whether the AS is a measure that can be usefully applied. A psychological measure is generally developed with specific purposes in mind, and using it generally leads to actions, decisions and consequences, either for a researcher, a clinician, a funding body, or a client (Maul *et al.*, 2013). Indeed, the AS was developed to impact counselling psychology research, and to be an outcome measure (Wood *et*

al., 2008). Moreover, I aim for my using it in this thesis to have consequences for whether or not it is employed as an outcome measure in the field of counselling psychology. Thus, pragmatism seems a natural and important epistemological position for psychological measurement as a whole, and for this thesis.

3.2.5. What is a psychological construct?

Pivotal to considering the epistemology of measurement in psychology, is a discussion of the ontology of the psychological constructs we seek to measure, often termed mental attributes, such as anxiety or authenticity.

The mental processes that make up mental attributes are formed of complex neural networks, a system made up of more than 100 billion neurones that can move through an almost infinite number of states (Fingelkurts & Fingelkurts, 2004). To make up mental attributes, these networks do not simply combine in additive or linear fashion, but assemble in non-linear dynamics that interact in multiplicative ways. In addition to this vastly complex neural system, mental processes are also affected by the external environment (Fingelkurts, Fingelkurts, & Neves, 2013). Thus, mental attributes clearly correspond to reality, as they are anchored in the neurophysiological, and thus physical and chemical, yet it is impossible to simply reduce them down to observable and measurable entities.

On the other hand, Guyon *et al.* (2018) hold that 'the reality of a mental attribute resides its functional appearance' (p. 4) and is thus 'a characteristic of the person as a social subject' (p. 5). In other words, the functional existence of authenticity depends on the presence of its manifestations that are observable in some way, socially, psychologically, or biologically. If someone shows no manifestation whatsoever of being authentic, then one cannot consider that person to be authentic. Thus, psychological attributes do not exist as phenomena independent of human perception or

interaction. They correspond to reality, in that they derive from neurophysiological processes in the brain, and yet they are also characteristics inferred from the behaviour of the person, and thus observer-dependent (Guyon *et al.*, 2018). They are, in other words, intersubjective realities, and they both fit into a purely realist epistemology, and yet they do not.

3.2.6. Pragmatism-realism

Based on this view of psychological constructs, Guyon and colleagues (2018) argue that adding pragmatism to realism, a pragmatism-realism epistemology, is the most coherent for psychological measurement, provided the realist position is not an extreme empiricist one that see reality as completely independent of human action or thought. Pragmatism-realism holds that our understanding of reality comes from our experience of the external world being organised by the layers of our conceptual and methodological theories, and our pragmatic concerns (Maul *et al.*, 2013). Putnam (1999) wrote of this epistemology: '*seeing* is always *seeing as*, and it is the interface between the world and the rich fabric of our concepts that jointly determines what we see.' (p. 20). This perspective sees psychological measurement as an empirical process seeking to gain knowledge about reality, but shaped by our pragmatic concerns.

Taking authenticity as an example, we can consider authenticity a social reality at this point in history, though it may be that in the future, some other concept will emerge and replace it. We talk of people being more or less authentic, based on the manifestations of our definition of authenticity. Moreover, quantifying and measuring authenticity can be useful, as I discussed in chapter 1. It may enable further research, or be useful to counselling psychologists, or to our clients, or to our wider community. Thus, from a pragmatism-realism perspective, using the AS to measure authenticity is reasonable, justifiable, and important.

3.2.7. Conclusion

In conclusion, I hold to a pragmatist-realist epistemological position in this thesis. This view fits with the concept of authenticity as being both a social construction, and yet corresponding to objective reality. From this epistemology, my experience and understanding of authenticity is organised and prioritised through my beliefs and hypotheses about it, as well as my methodological theories on how to approach it, and my pragmatic concerns and aims (Maul *et al.*, 2013). Taken together with authenticity being a social reality in our current society, I believe measuring it would be useful, which from a pragmatist-realist position, also suggests it would be meaningful to do so.

3.3. Aims, research questions and expected findings

The overall thesis has two aims, presented in section 1.5. Study 1 addressed the first of these aims: investigate the psychometric properties of the AS. Study 2 also addresses this aim, but in a clinical sample, which it is the first study to do, as every publication identified in Study 1 used non-clinical samples.

The emphasis in Study 2 was on addressing the second overall aim: investigate how the AS functions as an outcome measure.

3.3.1. Research questions

In Study 2, the first primary aim is addressed by the following research questions:

1. How reliable (internally consistent) is the AS in the clinical samples?
2. How does the AS relate to the CORE-10 at baseline, post-therapy and in change scores in the clinical samples (concurrent validity)?
3. Do the AS subscales intercorrelate as expected (construct validity)?

4. Is the factor structure of the AS replicable in the clinical samples (factorial validity)?

The second primary aim is addressed by the following research questions:

5. What are the completion rates for the AS?

6. How do clients rate the helpfulness of the AS on the TQHS, and how does this compare to the TQHS scores of the CORE-10?

7. Are there any floor or ceiling effects present in the AS scores?

8. How sensitive to change is the AS, and how does this sensitivity compare to that of the CORE-10?

3.3.2. Expected findings

1. I expect the internal consistencies for the AS and its subscales will be comparable to the estimates from the meta-analysis (AS: $\alpha = .84$, AL: $\alpha = .77$, AEI: $\alpha = .81$, SA: $\alpha = .82$).

2. I expect the estimated correlation between the AS and the CORE-10 at baseline to be moderate and negative, with 95% CIs that clearly show convergent validity for the AS (around $r = -.40$ for the full scale and each of the subscales).

3. I expect that the estimated correlation between the AS and the CORE-10 will be large post-therapy, (around $r = -.50$ for the full scale and each of the subscales), and that change scores on the two measures will share a moderate negative correlation (around $r = -.30$), with 95% CIs that exclude there being no relationship.

4. I expect to find the estimated subscale intercorrelations to indicate the three subscales are three related, but distinct, constructs. I also expect the estimated subscale intercorrelations, and the subscale correlations with the full AS, will be comparable to those found in the meta-analysis ($r = .34 - .44$ and $r = .64 - .82$, respectively).

5. I expect the AS will be sensitive to change, with AS scores significantly higher post-therapy than at baseline.

3.4. Design

Study 2 investigated the acceptability, reliability, validity and sensitivity to change of a pre-existing outcome measure, the Authenticity Scale (Wood *et al.*, 2008), in a clinical sample. It followed an observational longitudinal design, having participants complete psychometric outcome measures over the course of psychotherapy.

The length of the psychotherapy varied, between and within sites. Some participants only completed measures at baseline, whilst others completed measures all the way through therapy. The longest course of therapy in this study was 37 sessions. As the AS does not measure symptoms or occurrences, but rather stable traits that are hypothesised to change as a result of psychological intervention, participants did not complete the scale sessionally, but at set intervals. Participants also completed the CORE-10 (Barkham *et al.*, 2013), allowing me to compare how the AS functioned as an outcome measure in comparison to a measure that is already well established and widely used in this capacity ('CORE Users', n.d.). This data was then submitted to predetermined analyses specifically aimed at elucidating how suited the AS is as an outcome measure for psychotherapy.

3.5. Procedure

3.5.1. Ethical approval

This research project was submitted to the University of Roehampton Ethics Committee on August 4th, 2017 under the reference PSYC 17/275, and approved on September 18th, 2017

(Appendix D). A minor amendment to the client and therapist demographics forms was submitted on 27.11.17 and approved on 04.12.17 by the University of Roehampton Ethics Officer.

3.5.2. Sample size

I used G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) to calculate the approximate sample size I would need to complete the analyses I had planned. From the literature, I expected the AS and the CORE-10 to share a moderate correlation, with $r = -.40$. G*Power gave a minimum of 63 participants to detect this correlation as significant with power 95.2 %, one-tailed testing, and criterion alpha of .05. I used this planned correlation to calculate the sample size, rather than the t-tests looking at pre-post therapy comparisons in one sample, as these would require less statistical power than correlations between two measures. Thus, I aimed to recruit 63 participants. However, as more participants would add statistical power to the analyses, I decided to continue data collection until a predetermined cutoff point, even if I met the recruitment aim earlier. The cutoff point was October 31st 2018, as I estimated that was when I would have to start analysing the data in order to complete this project within the timeframe I had set. The freedom to continue collecting data was due to this thesis essentially being a pilot study: all the recruited sites had agreed to continue collecting data until the July 2019, with the aim of having sufficient data at that point to publish the study in a peer-reviewed journal.

3.5.3. Recruitment

Counselling services were recruited through opportunity sampling. I approached individual counselling and psychotherapy services with the study, aiming to indirectly recruit clients. This was done via an email sent to services on a database of placement providers for counselling psychology and psychotherapy students at the University of Roehampton. The email briefly

outlined the rationale and aims of the project (Appendix E). I first prioritised emailing services who affiliated themselves with humanistic approaches, as I reasoned these were likely to find the project interesting considering the strong Rogerian underpinnings of the AS (Wood *et al.*, 2008). Services connected to individuals on the research team were also approached, and the recruitment email was also posted on an online forum for person-centred research. There were no inclusion criteria pertaining to length of therapy, nor to therapy modality. Counselling services under the National Health Service (NHS) were not approach for recruitment, due to the additional requirements of submitting for NHS ethical approval.

Counselling services that were interested in participation were given the counselling service information sheet (Appendix F), and the services who agreed to participate, were given a Counselling Service Consent Form (Appendix G) to sign. They then put it to their individual therapists whether these wanted to take part or not, giving them the Therapist information sheet (Appendix H). In three cases, I attended a meeting with the service's therapists, explaining the project and answering questions. Therapists who agreed to participate signed a Therapist Consent Form (Appendix I), and completed a Therapist demographics form (Appendix J).

Participating therapists handed their new clients the Client Information Sheet and Client Consent Form (Appendices K & L) at the first point of meeting, in most cases an assessment session, and briefly introduced the project. Clients were asked to return the consent form signed for the first session should they wish to participate.

On the first therapy session, participating clients who had signed and returned the Client Consent Form were given the Client Demographics Form (Appendix M), and the first set of measures: the AS and the CORE-10. All completed forms were handed to the counselling service administrator or head to be stored securely in the service offices until I collected them.

3.6. Participants

3.6.1. Counselling services

Fifteen services, and two psychotherapists working privately, expressed interest in participating in this study. However, 11 of the services, and both individual therapists, withdrew their interest at various stages of the process. In most cases, this was due to the workload that would be required for participation. For eight of the services, the leadership teams wanted to take part, but withdrew their interest when their therapists did not want to. One service about to begin data collection withdrew after a change of staff, and three others were initially recruited, but stopped responding to emails.

In the end, four different sites were recruited to collect data for this study. Of these, one was a university research clinic from South England (referred to as ‘the Research Clinic’). This clinic was set up to offer free counselling to the community, whilst also contributing to research. It offers free counselling to adults (18+), and routinely collects outcome and process data, primarily involving asking clients to complete various measures when they come for their counselling sessions. The majority of the therapists at the clinic are current or recently qualified counselling, psychotherapy or counselling psychology postgraduate students.

The other three sites that were recruited were all independent counselling services, one in London, one in Sussex and one in the Midlands. These sites cover a broad demographic. The London site is a charity offering free counselling to anyone under 25 years old, including children. Clients over 18 are asked for a sessional donation, but this is not mandatory. The Coventry and Sussex sites are both commercial counselling services, though they offer an income-dependent sliding fee scale, and thus their clients come from a broad demographic.

There were no exclusion criteria for counselling modality or length of therapy, and all four sites offered counselling in several different psychotherapeutic approaches, and of various lengths.

The only inclusion criteria was being aged 18 years or older. In all cases, participation was voluntary, and was not compensated in any way.

3.6.2. Clients

Participants for this project were 67 clients receiving counselling or psychotherapy, from four different sites. See table 10 for client demographics.

Because of differences in demographics forms between sites, only 24 clients out of 67 were asked about their highest level of education. Of these, four (6.0%) had completed GCSE level secondary school (level 1), two (3.0%) had completed a post-secondary school qualification such as A-levels (level 2), seven (10.4%) had completed a vocational qualification (level 3), nine (13.4%) had completed an undergraduate degree (level 4), and two (3.0%) had completed a postgraduate degree (level 5).

Of the 67 participating clients, 18 (26.9%) only completed one Authenticity Scale, 20 (29.9%) completed two, 19 (28.4%) completed three, and five (7.5%) completed four, two (3.0%) completed five, and two (3.0%) completed six.

Table 10*Demographics of all clients who completed at least one measurement*

	Total (<i>N</i> = 67)	Research Clinic (<i>n</i> = 38)	Sussex (<i>n</i> = 21)	London (<i>n</i> = 3)	Midlands (<i>n</i> = 5)
Mean age in years (<i>SD</i>)	34.8 (12.1)	33.2 (2.0)	38.9 (2.9)	19.3 (1.3)	-
Missing	14 (20%)	9 (24%)	0	0	5 (100%)
Gender (<i>n</i> , %)					
Male	16 (24%)	6 (16%)	5 (24%)	1 (33%)	2 (40.0%)
Female	47 (70%)	21 (55%)	16 (76%)	2 (67%)	3 (60.0%)
Missing	4 (6%)	11 (29%)	0	0	0
Ethnic origin					
White British	39 (58%)	19 (50%)	19 (91%)	1 (33%)	-
White other	8 (12%)	7 (18%)	0	0	-
Asian	3 (5%)	2 (5%)	0	1 (33%)	-
Black	3 (5%)	3 (8%)	1 (5%)	0	-
Caribbean	2 (3%)	0	1 (5%)	1 (33%)	-
Missing	11 (16%)	0	0	0	5 (100%)

3.6.3. Therapists

Therapists from the Midlands service (*n* = 2) did not complete the demographics form, whereas therapists from the Clinic (*n* = 8) completed their service's own demographics form, which diverged somewhat from the one from this study.

There were 22 participating therapists in all, 16 (73%) female and four male (18%). Their mean age was 39.0 years (*SD* = 12.8), ranging from 24 years to 59. In terms of ethnicity, 17 (77%) were White, one was White Asian, and one was Black. Eight were from the Research Clinic, nine from the Sussex service, three from the London service and two from the Midlands service.

Table 11. shows therapists' professional identities, how long they have been qualified, and their principle therapeutic orientations.

Table 11
Therapist demographics

	Total (<i>N</i> = 22)
Mean age in years (<i>SD</i>)	39.0 (12.8)
Missing	3 (14%)
Currently a trainee	7 (32%)
Level of training	
5 (Diploma etc)	5 (23%)
6 (BA/BSc)	2 (9%)
7 (MA/MSc, PG Dip etc)	2 (9%)
8 (Doctoral degree)	9 (41%)
Missing	4 (18%)
Years qualified	
> 2	11 (50%)
> 6	4 (18%)
> 10	1 (5%)
≤ 10	2 (9%)
Missing	4 (18%)
Professional identity	
Counsellor	9 (41%)
Psychotherapist	0
Counselling Psychologist	9 (41%)
Missing	4 (18%)
Therapeutic orientation	
Humanistic	5 (23%)
Psychodynamic/analytic	1 (5%)
CBT	2 (9%)
Transactional analysis	1 (5%)
Integrative	5 (23%)
Pluralistic	4 (18%)
Missing	4 (18%)

3.7. Measures

3.7.1. The Authenticity Scale (AS)

The AS (Wood *et al.*, 2008) is a 12-item self-report measure, aiming to measure trait authenticity (Appendix A). Respondents rate how well each item describes them on a 7-point Likert scale, ranging from 1 ('Does not describe me at all') to 7 ('Describes me very well'), giving a total score between 12 and 84. The measure is made up of three 4-item subscales, *authentic living* (AL; e.g. item 1: 'I think it is better to be yourself, than to be popular. '), *accepting external influences* (AEI; e.g. item 5: 'I always feel I need to do what others expect me to do. '), and *self-alienation* (SA; e.g. item 2: 'I don't know how I really feel inside. '), with high scores on AL indicative of high authenticity, and high scores on the other two subscales indicative of low authenticity. For total AS scores, the AEI and SA subscales are reverse coded, before all three subscales are added together.

The copyright for the AS belongs to the American Psychological Association (APA). Upon application, I was given permission to reproduce the AS in this thesis (see Appendix B) and to use it in the study.

3.7.2. Clinical Outcomes in Routine Evaluation-10 (CORE-10)

The CORE-10 (Barkham *et al.*, 2013) is a 10-item self-report measure of general psychological distress (Appendix C), a short version of the 34-item CORE-OM (Evans *et al.*, 2000), with which it correlates at $r = .94$ (Barkham *et al.* 2013). It is not condition-specific or unifocal, but instead aims to measure common psychological issues, including depression, anxiety and trauma (e.g. item 1: 'I have felt tense, anxious or nervous', item 8: 'I have felt despairing or helpless', item 10: 'unwanted images or memories have been distressing me'), as well as levels of general and relational functioning (e.g. item 2: 'I have felt I have someone to turn to for support

when needed’, item 4: ‘Talking to people has felt too much for me’). It also includes one item measuring risk to oneself (item 6: ‘I have made plans to end my life’). On a 5-point Likert scale (0-4), respondents rate how often they have experienced what each item describes over the last week, from 'Not at all' to 'Most or all of the time', giving a total score between 0 and 40. Low scores are indicative of low psychological distress, and high scores indicate high levels of distress.

By including a wide range of items, the CORE-10 aims to capture the impact of an issue, such as depression, rather than merely the symptoms. Moreover, the CORE-10 is pantheoretical, thus aiming to be a useful standardised outcome measure across a wide range of settings and interventions (Barkham *et al.*, 2013). The CORE-10 is psychometrically sound, and Barkham *et al.* (2013) reported an internal reliability (alpha) of .90.

3.7.3. Therapy Questionnaire Helpfulness Survey (TQHS)

The TQHS (Di Malta *et al.*, unpublished) is a 10-item scale, aiming to measure clients’ experiences of using a specific measure. It is thus highly relevant when assessing the acceptability of a given measure. Specifically, the TQHS asks respondents to rate on a 5-point Likert scale (1-5) how much they agree or disagree with a series of statements. These relate to how helpful the measure was, how it made the respondent feel, how it affected the therapy, and whether it made the respondent think and reflect on themselves. For example, ‘I found this form helpful’ (item 1), ‘This form made me feel down’ (item 5) and ‘Using this form made the therapy better’ (item 8). The 10 items give a total score between 10 and 50, with higher scores indicating clients found the measure in question helpful, and lower scores indicating they found it unhelpful. The measure the TQHS is referring to is specified at the top of the TQHS.

As this measure is still under development, it is not publicly available, and not included in the appendix¹.

In our study, the TQHS was completed for both the AS (Wood *et al.*, 2008) and the CORE-10 (Barkham *et al.*, 2013), and had an internal reliability of $\alpha = .78$ for the AS, and $\alpha = .84$ for the CORE-10.

3.7.4. Measures completed at each service

Clients completed the AS and CORE-10 on the first and last sessions, as well as at set intervals, varying somewhat depending on what was feasible for the site. For the three independent counselling services, the interval was every 6th session (6th, 12th, 18th etc), whereas for the research clinic, clients completed the measures on the 4th and 10th sessions. The TQHS was completed on the 6th or 10th session, depending on the service. All participants completed the Authenticity Scale at least once. Participants from the Midlands service did not complete the TQHS, and participants from the Research Clinic did not complete the CORE-10. See table 12 for an overview of how many measures were completed at each recruited service.

¹ To view the TQHS, please email Professor Mick Cooper at mick.cooper@roehampton.ac.uk.

Table 12*Breakdown of how many of each measure was completed at each service*

Number of measures completed	Total (<i>N</i> = 67)	Research Clinic (<i>n</i> = 38)	Sussex (<i>n</i> = 21)	Midlands (<i>n</i> = 5)	London (<i>n</i> = 3)
AS					
1	18 (27%)	15 (39%)	3 (14%)	0	0
2	21 (31%)	10 (26%)	10 (48%)	1 (20%)	0
3	19 (28%)	6 (16%)	8 (38%)	2 (40%)	3 (100%)
4	5 (8%)	5 (13%)	0	0	0
5	2 (3%)	2 (5%)	0	0	0
6	2 (3%)	0	0	2 (40.0%)	0
Total AS (% of total)	159 (100%)	83 (52%)	47 (30%)	20 (13%)	9 (6%)
CORE-10					
1	5 (7%)	-	5 (24%)	0	0
2	11 (16%)	-	10 (48%)	1 (20%)	0
3	10 (15%)	-	6 (29%)	1 (20%)	3 (100%)
4	1 (1%)	-	0	1 (20%)	0
Total CORE-10 (% of total)	61 (100%)	0	43 (70%)	9 (15%)	9 (15%)
TQHS-AS					
Total TQHS-AS (% of total)	36 (100%)	20 (56%)	13 (36%)	-	3 (8%)
TQHS-CORE					
Total TQHS-CORE (% of total)	16 (100%)	-	13 (81%)	-	3 (19%)

3.8. Analyses

All analyses were decided upon a priori, so as to limit researcher bias. Analyses were carried out using SPSS (version 25; IBM, 2017). I used a criterion alpha level of .05 for all statistical tests. I used Cohen's (1988) benchmarks for the magnitude of correlation coefficients, *r*. According to these, .10 is a small effect size, .30 is a medium effect size, and .50 is a large effect size.

3.8.1. Estimation versus null-hypothesis testing

Null hypothesis testing (NHST) has traditionally been psychology's dominant approach to quantitative data analysis, but has been subject to cogent criticism, largely for not being informative enough (Cumming, Fidler, Kalinowski, Lai, 2012). Thus, the APA Publication Manual (2012; 6th edn.) recommends using the alternative paradigm, *estimation*. This essentially means asking the questions 'how much?' and 'to what extent?' rather the binary 'is there an effect?', which statistically translates into reporting effect sizes and confidence intervals where possible, rather than just probability tests (Cumming *et al.*, 2012). In this study, I report in accordance with the estimation paradigm where possible.

I used an online calculator (Lowry, 2019) to calculate the CIs for Pearson or Spearman correlations, as SPSS does not include this function.

Where appropriate, I also performed probability tests. These analyses were t-tests, correlations, one-way ANOVAs, confirmatory factor analyses (CFA) and multilevel modelling (MLM). For the vast majority of the correlational tests I had a clear hypothesis, based on the literature, as to the direction of the correlation, and thus most of these were one-tailed. Any two-tailed correlational tests have been labelled as such in the results sections.

3.8.2. Study-wise multiple-comparisons issue

In all I planned 85 comparisons and probability tests in Study 2, using the same dataset. This is important to bear in mind when considering the results of these tests, as each probability test increases the likelihood of a type I error, that is, of finding a significant result by chance. This likelihood is given by:

$$P(\text{at least one significant result}) = 1 - P(\text{no significant results})$$

Thus, there was a 98.7% study-wise likelihood of finding at least one significant result by chance. There are corrections available, for example the Bonferroni adjustment, but considering the small sample size, this would give a criterion alpha level so low that it is unlikely any of the tests would yield significant results, and all but ensure type II errors. In consultation with my supervisors, I decided to raise the issue here rather than apply any corrections to the criterion alpha level.

3.8.3. Sample sizes

Though the design of this research was for each client to complete the AS, CORE-10 and the TQHS, not all services were able to include all three measures. Though every client completed the AS at least once ($N = 67$), only 40 % completed the CORE-10 ($n = 27$), 54 % completed the TQHS for the AS ($n = 36$), and 24 % completed the TQHS for the CORE-10 ($n = 16$). Thus, sample sizes were considerably smaller than $N = 67$ for many of the analyses, limiting their statistical power. If I were basing this thesis on the NHST paradigm, I would have performed post hoc power analyses for all tests to explicate the statistical power for each. However, as I have largely based it on the estimation paradigm, reporting CIs, I have not done this. I reasoned that the wide CIs, coupled with low p-values, convey the power issues sufficiently, as argued by O’Keefe (2007). However, where results deviated markedly from my hypotheses, I calculated the observed (or post hoc) power for the analyses using G*Power (Faul *et al.*, 2009), based on population effect sizes from other studies as advocated by O’Keefe (2007).

3.8.4. Calculating baseline, post-therapy and change scores

For analyses that required only one timepoint, such as internal reliability (coefficient alpha), confirmatory factor analysis, subscale intercorrelations, and score distributions, I used the

baseline measurements, as this gave the most cross-sectional data. Because measurements took place at different points in therapy depending on the service and the client, I defined baseline as the first recorded measurement for each client. However, according to Baldwin, Berkeljon, Atkins, Olsen, and Nielsen (2009), who looked at the dose-effect rate of therapy, change is most rapid during the first eight sessions, and decreases as number of sessions increases. Thus, I needed to decide a cutoff point for what I considered baseline, that was both early enough in therapy to be reasonable in accordance with Baldwin *et al.* (2009), whilst including as much data as possible. I decide on the fifth session, aware that this was partly arbitrary, though informed by Baldwin *et al.* (2009). This cutoff point meant I excluded six cases, leaving $n = 61$ for the AS and $n = 27$ for the CORE-10.

I used the last measurements as post-therapy scores, which I compared to baseline to look at change. I used three sessions as a cutoff point, so any cases where there had been less than three sessions were excluded from the post-therapy scores, leaving $n = 45$ for the AS, and $n = 20$ for the CORE-10. This cutoff point was again based on Baldwin *et al.* (2009), who found that even after only three sessions, 35.8 % of clients who started therapy above the clinical cutoff for the Outcome Questionnaire-45 (OQ-45; Lambert, Hansen *et al.*, 1996) had experienced reliable and clinically significant improvement.

I calculated change scores for the AS and CORE-10, subtracting the baseline score from the post-therapy score, giving $n = 41$ for the AS, and $n = 20$ for the CORE-10.

3.8.5. Effects of client and therapist variables on AS scores

I assessed the effects of both participant and therapist demographic variables on AS scores. However, I decided these analyses were not sufficiently relevant to the aims of this study to warrant inclusion. Thus, they can be seen in Appendix P and Q.

I used analysis of variance (ANOVA) for the nominal variables (site, gender, sexuality, ethnicity and therapeutic orientation), Spearman's rank correlation coefficient (ρ) for the ordinal variables (level of education for participants and therapists, and therapist seniority) and Pearson product moment correlation coefficient for age, the one ratio level variable.

3.8.6. Acceptability

The first step in evaluating an outcome measure is looking at its acceptability, which refers to the participants' experience of completing a measure, essentially whether it is acceptable to them. One way to assess this is looking at completion rates, that is, whether participants actually complete the measure, or if certain items tend to be omitted. If certain items are omitted repeatedly it may point to a problem with that item, such as clients finding it hard to understand or upsetting. Completion rates were recorded for both the AS and CORE-10, and compared.

I also assessed acceptability by having participants complete the TQHS (Di Malta *et al.*, unpublished) for both the AS and CORE-10, which asks participants to rate their experience of the measure in question. I converted the TQHS scores to POMP scores so as to convey how the scores compared to the maximum possible score, and compared them for the AS and CORE-10.

3.8.7. Reliability

I assessed the internal reliability (coefficient alpha) of the full AS, as well as for each of the three subscales. The meta-analysis found an overall weighted alpha for the full scale of .83, and thus I expected the alpha for the full scale to be high. I also expected the internal consistency of each of the three subscales to be high, as the meta-analysis yield overall weighted alphas of .77 (AL), .81 (AEI) and .82 (SA).

I assessed the internal reliability of the CORE-10. Numerous studies have reported good internal reliability for the CORE-10 previously (e.g. Barkham *et al.*, 2013, $\alpha = .90$; O'Reilly, Peiper, O'Keeffe, Illback, & Clayton, 2016, $\alpha = .77$), and thus I expected similar alpha values.

I assessed the internal reliability of the TQHS for both the AS and CORE-10. No studies to date have reported on the reliability of the TQHS, but considering it is unidimensional and the items appear to measure a single construct, I expected high coefficient alpha values.

3.8.8. Validity

Validity refers to whether a scale measures what it aims to measure (DeVellis, 2012; see section 2.2.3.1. for a more thorough discussion). Unlike reliability, validity is a property of the scale rather than the scores, and thus validity reports for a measure found in the literature can be assumed to apply to the present study. As discussed in chapter 1, and assessed in chapter 2, there is a large amount of published data demonstrating the validity of the AS. Adding to this literature, I assessed the construct validity of the AS by both looking at its concurrent validity and how its subscales related to one another. I also assessed its factorial validity. The unique contribution of this study however, is that it is the first to assess the validity of the AS in a clinical population receiving psychotherapy.

3.8.8.1. Concurrent validity

By correlating the AS with the CORE-10, for both baseline, post-therapy scores, and change scores, I assessed the concurrent validity of the AS in the clinical samples. I expected the correlation coefficient to be moderate at baseline, as the CORE-10 is a measure of general psychological distress, thus conceptually different from the AS, yet still related (see chapters 1 and 2 for details). I expected it to be large post-therapy, as I expected CORE-10 scores to decrease and

AS scores to increase, thereby strengthening their correlation. I expected change on the AS and CORE-10 to be inversely correlated, but only moderately so, as they measure very different constructs.

3.8.8.2. Construct validity

As in Study 1 (see section 2.2.4.6.), I further explored the construct validity of the AS by looking at the intercorrelations of the three subscales, AL, AEI and SA, for both baseline and post-therapy scores, in light of their coefficient alphas. From the design model of the AS, I expected them to correlate moderately with each other, and to correlate highly with the full scale. These correlations, coupled with high coefficient alphas, would indicate that the data from Study 2 is consistent with the design model of the AS, supporting its validity.

Though this was the first study to use the AS with a clinical sample, I did not expect the AS subscales to correlate differently than in non-clinical samples. Thus, I expected correlations consistent with the meta-analysis, which found the subscale intercorrelations were all moderate ($r = .33 - .43$), and correlations with the full scale were all large ($r = .64 - .82$).

3.8.8.3. Factorial validity

I used confirmatory factor analysis (CFA) with the AMOS software (Arbuckle, 2017; version 25), to assess how well the three-factor model of the AS accounts for the variance in the AS baseline scores. Essentially, CFA is a statistical method for testing hypotheses about how different variables covary. Specifically, it allows me to draw a hypothetical model, and test how well my data fits this model (Hoyle, 2004). Thus, by drawing the three factor model that the AS is based on (Wood *et al.*, 2008), I can test if it is an appropriate model for the data in this study, potentially providing validity for the factor structure of the AS in a clinical sample.

Model fit was analysed using the chi-squared test of difference between the expected and the observed covariance matrices, the root-mean-square error of approximation (RMSEA), the standardized root-mean-square residual (SRMR), and the comparative fit index (CFI), as recommended by Hooper, Coughlan, and Mullen (2008). The chi-squared test can be problematic in assessing model fit, as it is very sensitive to sample size, and it is therefore recommended to also use the SRMR, RMSEA and CFI (Hooper *et al.*, 2008). There are several different recommendations for what the cutoffs for these fit indices should be (Hooper *et al.*, 2008). For the RMSEA, MacCallum, Browne, and Sugawara (1996) hold that a value below .08 indicates a good fit, whereas Hu and Bentler (1999) argue it should be .06 or below. For the SRMR, values below .08 are thought to be acceptable (Hu & Bentler, 1999), and so is a CFI value of .95 or above (Hu & Bentler, 1999), though a cutoff of .90 has been used previously (Hooper *et al.*, 2008).

As the three-factors of the AS theoretically combine to form one factor, authenticity, I also performed a CFA using a one-factor solution, and compared the fit indices of the one-factor and three-factor solutions to see which would best fit the clinical data

However, it is important to note that the statistical power and precision of CFA is sensitive to sample size, and Kline (2016) comments that as a general rule, sample sizes smaller than 100 generate untenable solutions. Marsh and Hau (1999) agree, holding that for a model with four indicators per latent factor, such as the model of the AS, a sample size larger than 100 is needed. Thus, though I expected the CFA to support the findings of Wood *et al.* (2008), showing that the three-factor model of the AS best fits the data, I was aware CFA might be problematic in this study, considering the AS baseline sample only had 61 participants.

3.8.9. Response distributions

3.8.9.1. Gaussian distribution

I tested whether the CORE-10 and AS baseline, post-therapy and change data had Gaussian distributions using the Shapiro-Wilk test (Shapiro & Wilk, 1965), as recommended by (Ghasemi & Zahediasl, 2012). This is important as parametric tests, such as t-tests, Pearson correlations, ANOVA and regression are all based on the assumption that data has a Gaussian distribution, and so if this is not the case, non-parametric tests should be used (Ghasemi & Zahediasl, 2012).

3.8.9.2. Floor and ceiling effects

The absence of floor and ceiling effects is important for any outcome measure, as any items subject to these effects will be limited in their ability to detect change at the upper or lower end of the measured construct (Lipsey, 1990; Hessling, Traxel, & Schmidt, 2004). Moreover, the limited variance that comes from floor and ceiling effects reduces the statistical power of correlation analyses, limiting the scale's usefulness in research (Hessling *et al.*, 2004).

I assessed for floor and ceiling effects for both scales as well as for the subscales of the AS, looking at response distributions for each item, and using POMP scores to look at the scoring distribution of each scale. However, POMP scores alone can only indicate such effects, and need to be seen in relation to *SDs*. As there is no agreed upon methodology for assessing floor and ceiling effects, I defined floor and ceiling effects at the scale level as a POMP score of below 10% or above 90%, respectively. At the item level, I defined floor and ceiling effects as > 29% of respondents using either the lowest or highest response category for an item, as suggested by Dean, Walker, and Jenkinson (2018).

3.8.10. Sensitivity to change

3.8.10.1. Estimating sensitivity

Sensitivity to change is key to any outcome measure, as it refers to the ability of a measure to pick up change that the client experiences. However, despite the importance of this characteristic, there is no single way in which to establish the sensitivity to change of a measure, nor is there consensus in the literature regarding what constitutes a sensitive measure (Husted, Cook, Farewell, & Gladman, 2000; Stratford & Riddle, 2005). Stratford and Riddle (2005) reviewed the outcome measure literature, specifically looking at methods for assessing measure responsiveness, and recommended using a simple paired t-test or repeated measures ANOVA to establish whether the means at two timepoints differ significantly. If there is a significant difference between baseline and post-therapy scores, it can be concluded the measure is sensitive to change (Husted *et al.*, 2000). However, t-tests focus solely on the statistical significance of change, which is useful, but only to an extent. This is because significance is a function of the magnitude of change and the variability of the measure, but also of the sample size, which clearly has nothing to do with the sensitivity of a measure (Husted *et al.*, 2000). Moreover, a t-test does not say how sensitive a measure is, just whether it is or not. Therefore, it is recommended to also report an effect size statistic of the change, a standardised response mean (SRM), given by the mean change divided by the *SD* of the change. Thus, a measure with a large degree of variability in change scores will have a large *SD* of change, and a low SRM (Stratford & Riddle, 2005). A higher SRM score indicates greater sensitivity to change.

In accordance with Stratford and Riddle (2005), I used a paired-samples t-test to compare baseline and post-therapy AS and CORE-10 scores, before calculating the effect size of the change, SRM scores. I expected the AS to be sensitive to change, and for its SRM score to be comparable to that of the CORE-10.

3.8.10.2. Reliable and clinically significant change

Reliable change (RC) and clinically significant change (CSC) are concepts introduced by Jacobson, Follette, and Revenstorf (1984), and differ from other methods of change analyses in that they look at individual change rather than group means. Though they are often used to assess the quality of an intervention or of a service, it is important to ascertain whether the AS is sensitive enough to not only pick up change, but RC and CSC.

Reliable change (RC). RC refers to whether the change on a given measure is sufficiently great that it is unlikely it is simply due to the measurement error (Evans, Margison, & Barkham, 1998). In other words, if a participant's scores on the AS are higher post-therapy than at baseline, is the magnitude of this change beyond the range that could be attributed to the variability of the measure itself? Assessing whether participants have achieved reliable change involves calculating a reliable change criterion (RCrit), which is a function of the standard deviation of the initial measurement, and the coefficient alpha of the measure (see Evans *et al.* 1998). Traditionally, RCrit uses the 95% false positive criterion (Evans *et al.*, 1998). However, as the RCrit is naturally higher for measures with greater item breadth, such as the AS, as compared to measures with a tight focus, such as the PHQ-9, Barkham *et al.* (2013) recommend using a 90 % criterion instead. This means that if the change on a measure is greater than the calculated RCrit, there is a 90 % likelihood that this change was due to actual improvement, rather than the unreliability of the measure (Evans *et al.*, 1998).

I calculated the reliable change criterion (RCrit) for each scale using a 90% criterion, and the proportion of participants who showed reliable improvement and deterioration on each measure. For the AS, the RCrit was 11.5, and for the CORE-10 it was 6.4.

Clinically significant change (CSC). The second way of considering individual change is looking at whether it is clinically significant. CSC refers to whether a client's scores have moved from a 'dysfunctional' scoring range to a 'normal' scoring range (Evans *et al.*, 1998). Thus, the important consideration is what is a typical score for a 'dysfunctional', or clinical, population, and what is a typical score for a 'normal', or non-clinical, population. Once these benchmarks are defined, we can calculate the point which clients must move beyond in order to qualify as having achieved CSC. Importantly, for change to be clinically significant, it must also be reliable, which means RC is a prerequisite for CSC. Jacobson *et al.* (1984) suggest three different criteria for determining whether CSC has taken place. Depending on the data, either of these may be the most appropriate. In the case of this study, I expected the distribution of non-clinical and clinical scores to overlap, in which case Evans *et al.* (1998) recommend using 'method C' (Jacobson *et al.*, 1998), which takes into account both a normative population score and a clinical population score and calculates a cutoff point where the likelihood of being in the non-clinical population is equal to being in the clinical population. This cutoff is given by:

$$\frac{(mean_{clin} \times SD_{norm}) + (mean_{norm} \times SD_{clin})}{SD_{norm} + SD_{clin}}$$

In this study, I used the baseline mean AS score as the clinical population score, and selected a study from the meta-analysis for the mean non-clinical population score. I needed a study with a large sample of non-clinical participants that reported full AS scores and *SDs*, not just subscale scores. The study that best fit these criteria was Boyraz and Kuhl (2015). Their sample consisted of 619 adults recruited from two websites that publish links to psychological studies,

with a mean age of 26.9 years ($SD = 10.5$). The mean item AS score in this sample was 5.12 ($SD = 1.06$)

As I did not have access to CORE-10 non-clinical population data, I could not calculate the criterion for determining CSC for the CORE-10. Instead, I used 11, the CSC criterion reported by Barkham *et al.* (2013) for the CORE-10. Thus, clients who scored above 11 at baseline and 11 or below at post-therapy, would have showed CSC on the CORE-10.

I calculated the percentage of clients who showed clinically significant improvement and deterioration on both the AS and the CORE-10.

3.8.10.3. Comparing the sensitivity of the AS and CORE-10

I compared the sensitivity to change of the AS and the CORE-10. I reasoned that as the CORE-10 is already a widely used outcome measure (see section 1.4.3.2.), it would provide support for the AS as an outcome measure if its sensitivity was comparable to that of the CORE-10.

To test whether the change differed significantly between the measures, I standardised both sets of scores to POMP scores, as the CORE-10 and AS use different scoring scales (0-4 and 1-7 respectively). I calculated POMP change scores by subtracting baseline POMP scores from post-therapy POMP scores. Because positive change on the CORE-10 means lower scores, I subtracted the CORE-10 POMP change score from 0, making both POMP change scores positive. I then used a t-test to examine whether the change POMP scores differed significantly between the measures.

3.8.10.4. Multilevel modelling

Finally, I examined sensitivity to change using MLM, a statistical method that extends conventional regression methods, and is well suited for analysing hierarchical data (Gill &

Womack, 2013). It essentially performs a regression analysis on all the data, and gives the slope of the model, that is, how much the average participants' scores change per session, as well as how well that linear slope fits the data. The model can be constrained to a fixed linear fit, meaning it assumes the relationship between number of sessions and AS scores was the same for all participants, or it can allow change to be non-linear, which means it first calculates the linear model, before calculating how much each participant deviates from that model. Similarly to CFA, the question is then which model best fits the data

The strength of MLM over a t-test is that, rather than just comparing averaged pre- and post-scores, data can be individualised, that is, the MLM can include every data point for every client, regardless of how many sessions each client has had, or if the measurements occurred on the same sessions (Tasca & Gallop, 2009). Indeed, the t-test comparing baseline and post-therapy scores only uses 67% of the data gathered, limiting the statistical power available. Furthermore, MLM avoids the problem of the somewhat arbitrary cutoff points I used to define baseline and post-therapy scores, circumventing that source of error completely. MLM is thus well suited for psychotherapy research in naturalistic settings such as this, where number of sessions can vary greatly between clients.

I had planned to perform the MLM in SPSS, but after both my supervisor and I had spent a considerable amount of time attempting it, I decided it would be better to perform it with the R software, and the nlme add-on package (Pinheiro, DebRoy, Sarkar, & R Core Team, 2018). Thus, though I planned the analyses and interpreted the results, the actual MLM analysis was performed by my supervisor, Professor Chris Evans, in consultation with myself, the main author, and I take responsibility for the analysis.

We tested three different models and compared which fit the data best. The first model was simply to provide a point of comparison, and assumed that mean AS scores across participants

were the same, ignoring session numbers completely. The second model allowed each participant to have a different intercept point with the y-axis (session number = 0), and assumed a fixed linear model, that is, that every participant shows the same linear relationship between session number and AS scores. The third model allowed each participant to have a different relationship between session number and AS scores.

Model fit was assessed with the Akaike Information Criterion (AIC; Akaike, 1973) and the Schwarz's Bayesian Criterion (BIC; Schwarz, 1978). Lower values for AIC and BIC indicate that more variance is explained, and thus better fit. ANOVA was used to test whether the models differed significantly.

Like with the t-test analysis, I expected the MLM to find the AS was sensitive to change.

Chapter 5. Results

5.1. Acceptability

5.1.1. Completion rates

In all, 159 AS measures were collected. Of these, four measures had omitted items, though never more than one. This gave a completion rate of 99.97% for the AS.

In all, 61 CORE-10 measures were completed. Of these, one had an omitted item, giving a completion rate of 99.98%.

5.1.2. Therapy Questionnaire Helpfulness Survey (TQHS) scores

The AS had a mean TQHS score of 30.8 ($SD = 5.7$; $n = 36$), which is a POMP score of 52%. The CORE-10 had a mean TQHS score of 31.3 ($SD = 5.5$; $n = 14$), which is a POMP score of 53%. The mean AS and CORE-10 TQHS scores were not significantly different: $t(12) = -.91$, $p = .38$. The mean difference was $-.69$, 95% CI $[-2.36, 0.97]$.

5.2. Reliability

5.2.1. Internal reliability (coefficient alpha)

At baseline, the internal reliability of the AS was $\alpha = .80$, 95% CI $[.71, .87]$. For the AI subscale it was: $\alpha = .81$, 95% CI $[.71, .88]$; for the AEI: $\alpha = .86$, 95% CI $[.80, .91]$; and for the SA, $\alpha = .78$, 95% CI $[.67, .86]$. These coefficient alpha values were all comparable with those found by the reliability generalisation in the meta-analysis (see table 13).

Table 13*Coefficient alphas at baseline and post-therapy compared with the meta-analysis*

Scale	Baseline (α)			Post-therapy (α)			Meta-analysis (α)		
	Est.	ICL	uCL	Est.	ICL	uCL	Est.	ICL	uCL
<i>Authenticity scale</i>	.80	[.71	.87]	.88	[.83	.93]	.84	[.82	.86]
<i>Authentic Living</i>	.81	[.71	.88]	.84	[.75	.90]	.77	[.75	.79]
<i>Accepting External Influences</i>	.86	[.80	.91]	.89	[.83	.93]	.81	[.80	.83]
<i>Self-Alienation</i>	.78	[.67	.86]	.88	[.82	.93]	.82	[.80	.85]

Note: **Est** = estimate; **ICL** = lower confidence limit; **uCL** = upper confidence limit.

At baseline, $n = 61$. Post-therapy, $n = 45$. Meta-analysis, $n = 8,823$.

At baseline the internal reliability of the CORE-10 was $\alpha = .86$, 95% CI [.76, .93] ($n = 27$).

The TQHS for the AS had an internal reliability of $\alpha = .78$, 95% CI [.65, .87], and for the CORE-10, $\alpha = .84$, 95% CI [.68, .94].

5.3. Validity

5.3.1. Concurrent validity

See table 15 for baseline and post-therapy scale correlations.

Baseline. At baseline, the AS and CORE-10 shared a significant, moderate negative correlation: $r(25) = -.36$, $p = .032$, 95% CI [-.65, .02].

At baseline, the AL subscale did not significantly correlate with the CORE-10, $r(25) = -.01$, $p = .47$, 95% CI [-.39, .37]. The AEI subscale shared a significant, moderate positive

correlation with the CORE-10: $r(25) = .39, p = .022, 95\% \text{ CI } [.01, .67]$, and so did the SA subscale: $r(25) = .34, p = .041, 95\% \text{ CI } [-.05, .64]$.

At baseline, the AS shared a significant, large positive correlation with its AL subscale: $r(59) = .65, p < .001, 95\% \text{ CI } [.48, .77]$, and significant, large negative correlations with its AEI, $r(59) = -.79, p < .001, 95\% \text{ CI } [-.87, -.67]$, and SA subscales, $r(59) = -.65, p < .001, 95\% \text{ CI } [-.77, -.48]$.

Post-therapy. At post-therapy, the AS and CORE-10 shared a significant, large negative correlation: $r(17) = -.73, p < .001, 95\% \text{ CI } [-.89, -.41]$.

At post-therapy, the AL subscale did not significantly correlate with the CORE-10, $r(18) = -.35, p = .064, 95\% \text{ CI } [-.69, .11]$. The AEI subscale shared a significant, moderate positive correlation with the CORE-10: $r(18) = .43, p = .028, 95\% \text{ CI } [-.02, .73]$, and the SA subscale shared a large positive correlation: $r(17) = .78, p < .001, 95\% \text{ CI } [-.51, .91]$.

At post-therapy, the AS shared a significant, large positive correlation with its AL subscale: $r(43) = .72, p < .001, 95\% \text{ CI } [.54, .84]$, and significant, large negative correlations with its AEI, $r(43) = -.79, p < .001, 95\% \text{ CI } [-.88, -.65]$, and SA subscales, $r(43) = .83, p < .001, 95\% \text{ CI } [.71, .90]$.

Change. Correlation analysis revealed that change on the AS and change on the CORE-10 were non-significantly correlated: $r(17) = -.35, p = .070, 95\% \text{ CI } [-.69, .12]$.

A post hoc power analysis revealed that a sample size of 49 would be required for a one-tailed analysis to detect an effect of $r = -.35$ with power of .80.

Table 14 shows the correlations between change on all scales.

Table 14*Matrix of intercorrelations of change for all scales*

Scale	n	Change (post-therapy - baseline)				
		CORE-10	AS	AL	AEI	SA
CORE-10	20	-	-.35 <i>ns</i>	-.30 <i>ns</i>	.03 <i>ns</i>	-.44*
Authenticity Scale	41		-	.45**	.74**	.88**
Authentic Living	43			-	-.01 <i>ns</i>	.12 <i>ns</i>
Accepting External Influences (R)	42				-	.53**
Self-Alienation (R)	42					-

Note: **AS:** Authenticity Scale; **AL:** Authentic Living subscale; **AEI:** Accepting External Influences subscale; **SA:** Self-Alienation subscale. **R:** Reverse coded

5.3.2. Construct validity (subscale intercorrelations)

See table 15 for a matrix of the intercorrelations of the subscales.

Baseline. At baseline, the AL subscale did not significantly correlate with the SA subscale: $r(59) = -.04, p = .39, 95\% \text{ CI } [-.29, .21]$. The AL did significantly correlate with the AEI subscale: $r(59) = -.40, p < .001, 95\% \text{ CI } [-.59, -.17]$. The AEI and SA subscales correlated significantly at baseline: $r(59) = .23, p = .04, 95\% \text{ CI } [-.02, .46]$.

In the meta-analysis, the combined population effect for the correlation between AL and SA was: $r(5,673) = -.36 (p < .001, 95\% \text{ CI } [-.41, -.32])$. As the baseline correlation between the AL and SA subscales in Study 2 was non-significant, I calculated the likelihood of finding a significant result assuming the observed sample effect was exactly the same as the estimate population effect from the meta-analysis. This gave a power of .89.

Post-therapy. Post-therapy, the AL subscale correlated significantly with the SA subscale: $r(44) = -.44, p < .001, 95\% \text{ CI } [-.65, -.17]$; and with the AEI subscale: $r(44) = -.29, p = .03, 95\% \text{ CI } [-.54, 0]$. The AEI and SA subscales correlated significantly post-therapy: $r(43) = .49, p < .001, 95\% \text{ CI } [.23, .69]$.

Table 15
Matrix of intercorrelations for the AS, its subscales and the CORE-10

Scale	Baseline					Post-therapy				
	CORE-10	AS	AL	AEI	SA	CORE-10	AS	AL	AEI	SA
CORE-10	-	-.36*	-.01 <i>ns</i>	.39*	.34*	-	-.73**	-.35 <i>ns</i>	.43*	.78**
AS		-	.65**	-.79**	-.64**		-	.72**	-.79**	-.83**
AL			-	-.40**	-.04 <i>ns</i>			-	-.29*	-.44**
AEI				-	.23*				-	.49**
SA					-					-

Note: **AS:** Authenticity Scale; **AL:** Authentic Living subscale; **AEI:** Accepting External Influences subscale; **SA:** Self-Alienation subscale.

At *baseline*, $n = 61$ for the AS and its subscales; $n = 27$ for the CORE-10. At *post-therapy*, $n = 43$ for the AS and its subscales; $n = 20$ for the CORE-10.

* $p < .05$ ** $p < .001$. *ns* = non-significant

5.3.3. Factorial validity.

I performed a CFA with the baseline AS scores, using the maximum likelihood model of estimation. I tested a three-factor solution consistent with the theoretical model of the AS, which, according to lenient cutoffs (Hooper *et al.*, 2008) provided a reasonable fit for the data: $\chi^2(51) = 68.83 (p = .049)$, CFI = .94, RMSEA = .08 (90% CI = .01 - .12), SRMR = .08.

However, when inspecting the standardised regression weights, it was clear the data did not fit a hierarchical model at all, as the regression weight between the AEI and the full AS was -2.00, an impossibility, and the loading of the AEI error term was 3.99 (see figure 10). This is called a Heywood case, and indicates that the converged solution is inadmissible (Kline, 2016).

According to Kline (2016), Heywood cases can be due to small sample sizes ($N < 100$), which, considering the sample size in this analysis was 61, is a likely source of the issue.

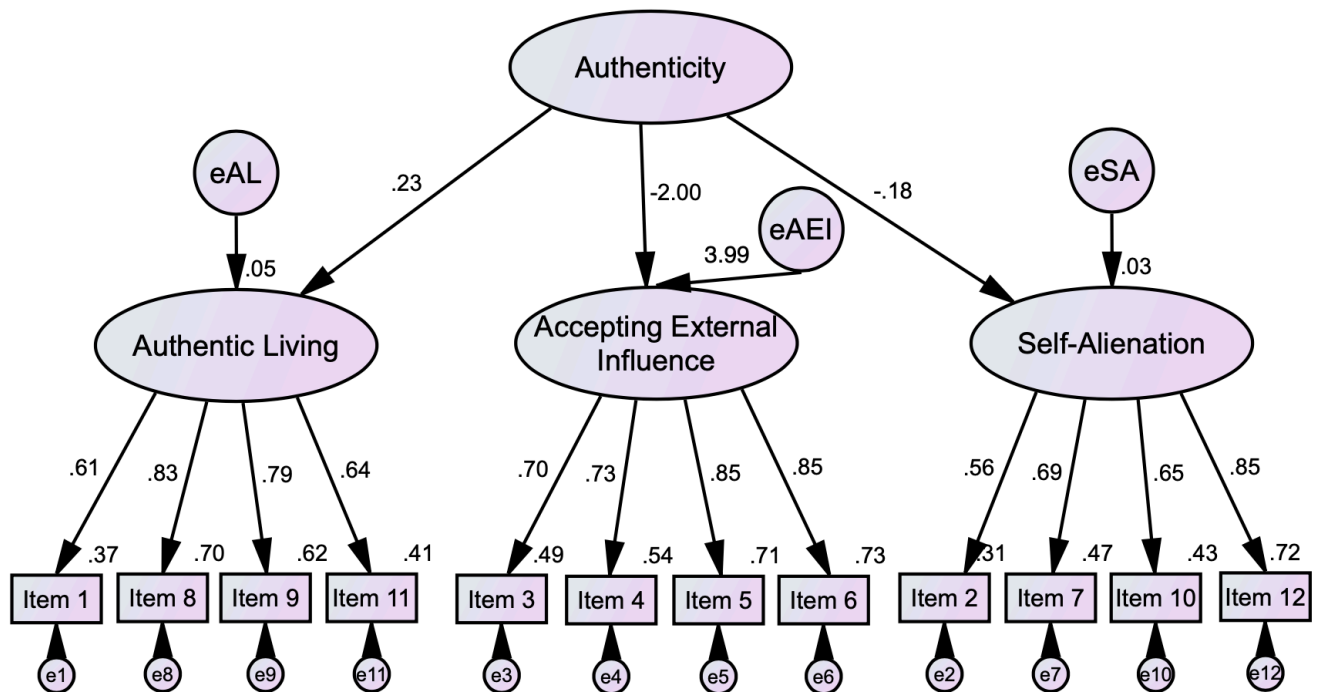


Figure 10. Hierarchical model of the AS with standardised regression weights.

5.4. Response distributions

5.4.1. Gaussian distribution

I examined the distribution shape of the AS and CORE-10 baseline, post-therapy, and change data. The Shapiro-Wilk test was non-significant at both baseline ($S-W = .96$, $df = 61$, $p = .07$) and post-therapy ($S-W = .98$, $df = 45$, $p = .55$), as well as for change scores ($S-W = .94$, $df = 19$, $p = .25$), indicating the data had a Gaussian distribution. See figures 11 - 13 for histograms of AS baseline, post-therapy and change scores, with Gaussian distribution curves.

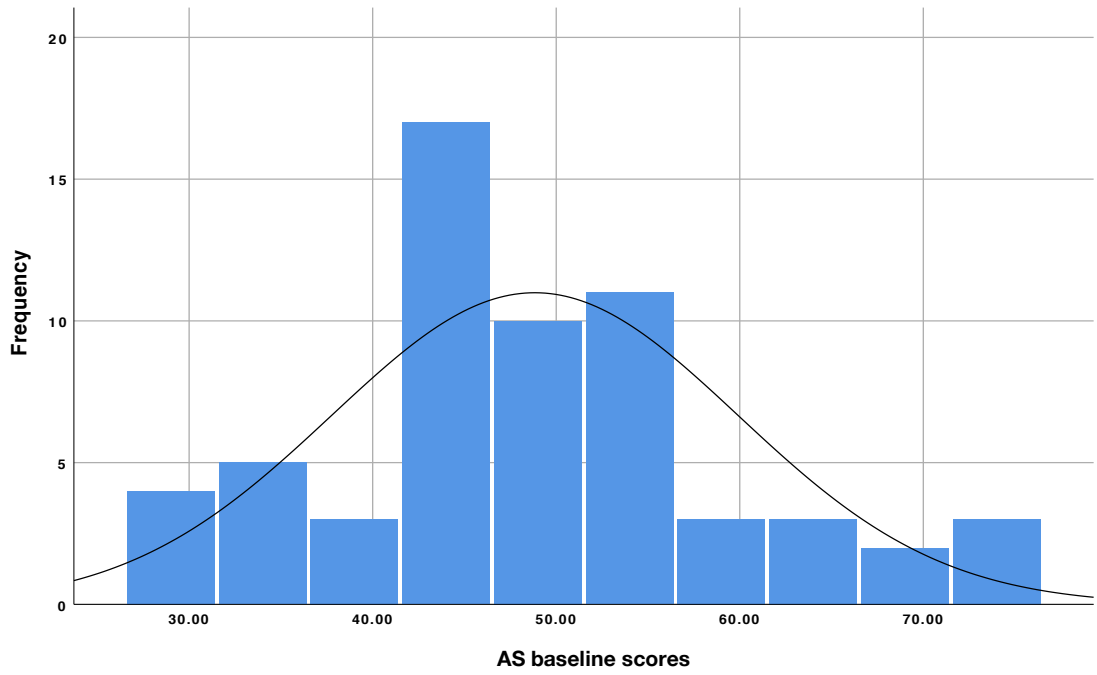


Figure 11. Histogram of AS baseline scores with Gaussian curve.

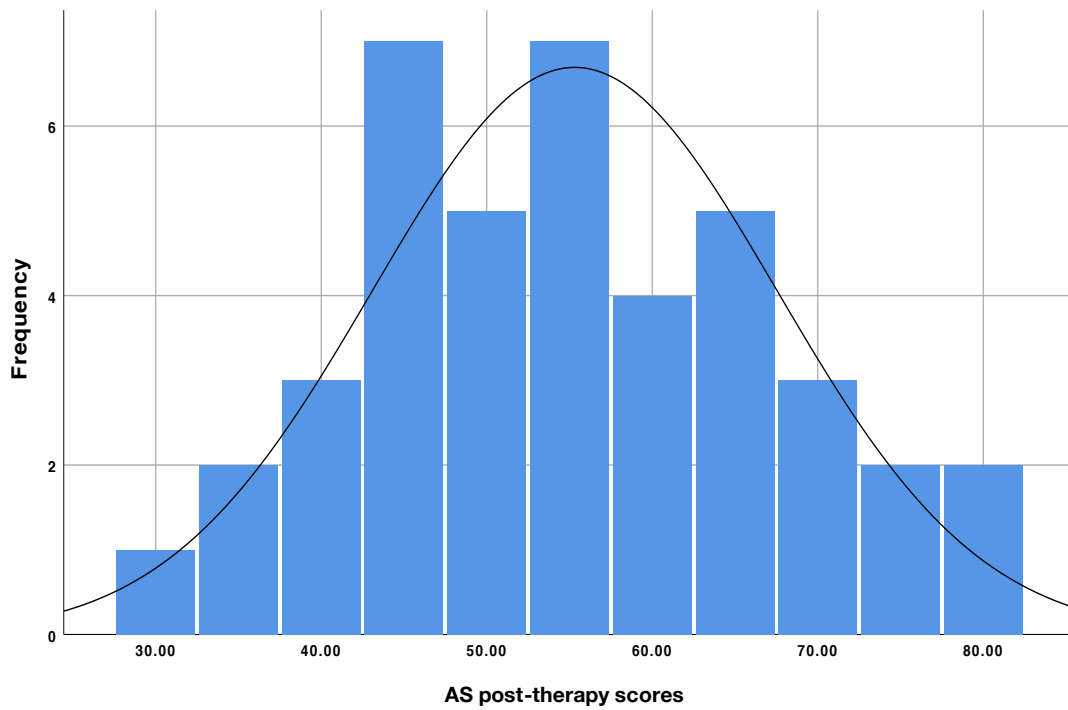


Figure 12. Histogram of AS post-therapy scores with Gaussian curve.

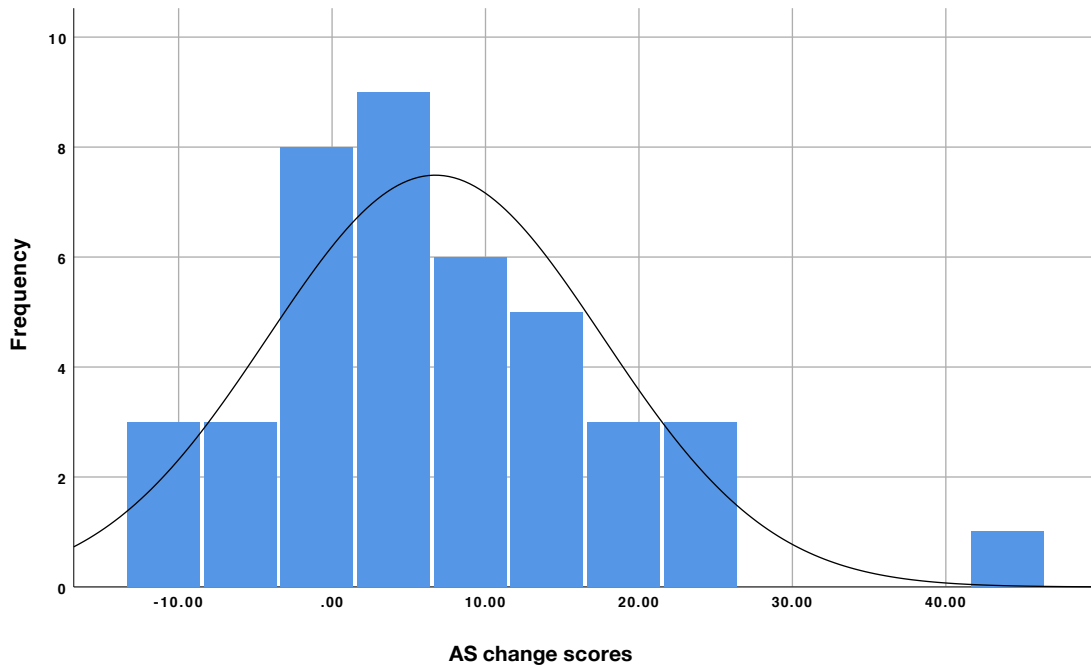


Figure 13. Histogram of AS change scores with Gaussian curve.

For the CORE-10, the Shapiro-Wilk test was non-significant at both baseline ($S-W = .98$, $df = 27$, $p = .82$) and post-therapy ($S-W = .94$, $df = 16$, $p = .40$), as well as for change scores ($S-W = .96$, $df = 19$, $p = .66$), indicating the data had a Gaussian distribution.

5.4.2. Floor/ceiling effects

At baseline, no floor (POMP < 10%) or ceiling (POMP > 90%) effects were present in AS scores, nor in any of its subscales. See table 16 for an overview of the AS and its subscales.

At baseline, the CORE-10 did not exhibit any floor or ceiling effects at the scale level ($M = 16.70$, $SD = 7.41$), though its POMP score of 41.8% was lower than the AS.

Table 16*Means and POMP scores for the AS and its subscales*

Scale	Baseline				Post-therapy			
	<i>n</i>	Mean	SD	POMP %	<i>n</i>	Mean	SD	POMP %
<i>Authenticity Scale</i>	61	48.80	11.07	51.1 %	45	55.58	12.49	60.5%
<i>Authentic Living</i>	61	19.69	4.83	65.4 %	47	19.91	4.88	66.3 %
<i>Accepting External Influences (R)</i>	61	14.56	5.43	44.0%	46	16.09	5.60	50.4 %
<i>Self-Alienation (R)</i>	61	14.59	5.67	44.1 %	46	19.71	5.64	65.5 %

Note: **R**: Reverse coded

No items on the AS at baseline exhibited floor or ceiling effects (see table 17): participants used the entire scoring range for each item, except for items 8, 9, and 11, where the lowest score was not used.

Table 17*Item completeness and response distribution for the AS (Baseline)*

Item	Scale	<i>n</i>	Mean	SD	POMP %	Response category (%)						
						Does not describe me at all	2	3	4	5	6	Describes me very well
1	AL	67	4.94	1.70	65.7 %	3.0	7.5	9.0	19.4	19.4	17.9	23.9
2R	SA	67	3.85	1.88	47.5 %	11.9	14.9	20.9	14.9	13.4	13.4	10.4
3R	AEI	67	3.49	1.49	41.5 %	10.4	16.4	23.9	22.4	17.9	7.5	1.5
4R	AEI	67	4.31	1.77	55.2 %	6.0	13.4	11.9	22.4	17.9	14.9	13.4
5R	AEI	67	3.43	1.66	40.5 %	9.0	25.4	20.9	25.4	4.5	7.5	7.5
6R	AEI	67	3.48	1.64	41.3 %	11.9	14.9	31.3	16.4	10.4	10.4	4.5
7R	SA	67	3.70	1.83	45.0 %	13.4	19.4	10.4	32.9	13.4	11.9	7.5
8	AL	67	5.13	1.49	68.8 %	0.0	4.5	9.0	23.9	19.4	17.9	25.4
9	AL	67	4.79	1.46	63.2 %	0.0	7.5	13.4	19.4	25.4	20.9	13.4
10R	SA	67	3.09	1.87	34.8 %	25.4	22.4	14.9	13.4	7.5	11.9	4.5
11	AL	67	4.96	1.47	66.0 %	0.0	6.0	10.4	22.4	23.9	17.9	19.4
12R	SA	67	3.82	1.82	47.0 %	10.4	16.4	20.9	14.9	17.9	9.0	10.4

Note: **AS**: Authenticity Scale; **AL**: Authentic Living subscale; **AEI**: Accepting External Influences subscale; **SA**: Self-Alienation subscale. **R**: reverse coded

5.5. Sensitivity to change

5.5.1. Baseline and post-therapy comparison

The AS and its subscales. A paired-samples t-test revealed there was a significant difference between mean baseline and post-therapy AS scores, $t(40) = 3.95, p < .001$, with a mean difference of 6.73, 95% CI [3.28, 10.18]. Participant AS scores were significantly higher after therapy.

Paired-samples t-tests revealed there was a significant difference between mean baseline and post-therapy AEI scores, $t(41) = 3.22, p = .003$, with a mean difference of -2.21, 95% CI [-3.61, -0.82]. There was also a significant difference between mean baseline and post-therapy SA scores, $t(41) = 4.47, p < .001$, with a mean difference of -4.64, 95% CI [-6.74, -2.54]. The difference between mean baseline and post-therapy AL scores was non-significant, $t(42) = 0.19, p = .85$, with a mean difference of 0.12, 95% CI [-1.09, 1.33]. Participant scores on the AEI and SA subscales were significantly higher after therapy, whilst AL scores did not change significantly.

The CORE-10. There was a significant difference between mean baseline and post-therapy CORE-10 scores, $t(19) = -3.36, p = .003$, with a mean difference of -6.15, 95% CI [-9.98, -2.32]. Participant CORE-10 scores were significantly lower after therapy.

5.5.2. Effect size of change

I calculated change scores by subtracting the baseline scores from the post-therapy scores. I calculated SRM scores for the AS, its subscales, and the CORE-10 (see section 3.8.10.1). The AS had an SRM of 0.62, the AL, AEI, and SA subscales had SRMs of 0.03, 0.50, and 0.69, respectively. The CORE-10 had an SRM of 0.75.

In the 19 clients who had change scores on both the AS and CORE-10, the SRM of the AS was 0.73, and 0.77 for the CORE-10.

5.5.3. POMP change scores

In the 19 clients with change scores on both measures, the mean AS change POMP score was 12.8%, and the CORE-10 change POMP score was -16.1%. On average, participants scored 13% higher on the AS after therapy, and 16% lower on the CORE-10. See Table 18 for an overview of mean and POMP change scores for each scale, as well as SRM effect sizes.

A t-test revealed the difference between AS change POMP scores and CORE-10 change POMP scores was non-significant, $t(18) = -.65, p = .53$, with a mean difference of -3.3, 95% [-13.9, 7.3].

Table 18

Descriptive statistics for change on the AS, its subscales and the CORE-10.

Scale	Change				
	<i>n</i>	Mean	SD	POMP %	SRM
<i>Authenticity Scale</i>	41	6.73	10.92	9.4 %	0.62
<i>Authentic Living</i>	43	0.12	3.94	0.5 %	0.03
<i>Accepting External Influences (R)</i>	42	2.21	4.45	9.4 %	0.50
<i>Self-Alienation (R)</i>	42	4.64	6.67	19.4 %	0.69
<i>n = 19</i>					
<i>Authenticity Scale</i>	19	9.21	12.60	12.8 %	0.73
<i>CORE-10</i>	19	-6.42	8.32	-16.1 %	0.77

Note: The two bottom rows of the table only include data from the 19 clients with change scores on both the AS and CORE-10. **R** = reverse coded.

5.5.4. Clinically significant change (CSC) and reliable change (RC)

I calculated the criterion for CSC using the formula suggests by Jacobsen *et al.* (method C; 1984; see section 3.8.10.2). This gave an AS mean item criterion of 4.56, and a scale criterion of 54.7. Thus, any clients who had an AS score below 54 at baseline, and above 55 at post-therapy (or vice versa), and whose change was reliable, would have showed CSC. For the CORE-10, I used 11 as the CSC criterion, the same as reported by Barkham *et al.* (2013).

The RCrit for the AS was 11.5, and for the CORE-10 it was 6.4. See Table 19 for an overview of how many clients showed RC and CSC on the AS and CORE-10.

Table 19

Percentage of clients showing RC and CSC on the AS and CORE-10

	Reliable change			Total
	Reliable improvement	No reliable change	Reliable deterioration	
<i>Authenticity Scale (n = 41)</i>				
Clinically significant improvement	10 (24%)	2 (5%)	0 (0%)	12 (29%)
No clinically significant change	2 (5%)	27 (66%)	0 (0%)	29 (71%)
Clinically significant deterioration	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total	12 (29%)	29 (71%)	0 (0%)	41 (100%)
<i>Authenticity Scale (n = 19)</i>				
Clinically significant improvement	6 (32%)	1 (5%)	0 (0%)	7 (37%)
No clinically significant change	1 (5%)	10 (53%)	0 (0%)	11 (58%)
Clinically significant deterioration	0 (0%)	1 (5%)	0 (0%)	1 (5%)
Total	7 (37%)	12 (63%)	0 (0%)	19 (100%)
<i>CORE-10 (n = 19)</i>				
Clinically significant improvement	9 (45%)	1 (5%)	0 (0%)	10 (50%)
No clinically significant change	0 (0%)	7 (37%)	0 (0%)	7 (37%)
Clinically significant deterioration	0 (0%)	1 (5%)	1 (5%)	2 (11%)
Total	9 (47%)	9 (47%)	1 (5%)	19 (100%)

Note: The bottom two tables ($n = 19$) show the percentages of clients, who had both the AS and CORE-10 change scores, showing RC and CSC.

5.5.5. Multi-level modelling

I tested three different models for best fit to the AS data (see table 20).

Table 20

Estimates of multilevel models for AS scores against number of sessions.

Model	Coefficient	SE	t	df	p	Model fit	
						AIC	BIC
Null model						1106.1	1115.0
Intercept	51.48	1.28	40.29	80	< .001		
Fixed linear slope						1085.8	1097.7
Intercept	48.78	1.41	34.62	79	< .001		
Slope	0.46	0.089	5.16	79	< .001		
Free linear slope						1088.0	1106.0
Intercept	48.53	1.38	35.26	79	< .001		
Slope	0.52	0.10	5.01	79	< .001		

N = 67

Observations = 147

ANOVA revealed the AIC for the null model and for the fixed linear model were significantly different: $F(3, 4) = 22.32, p < .001$. The fixed linear slope model fit the data significantly better than the null model.

The AIC for the fixed linear and the free linear models were not significantly different: $F(4, 6) = 1.72, p = .42$.

As the fixed and free linear models explained the same proportion of variance, I used the fixed linear model, as it was the most parsimonious. This gave a slope of 0.46, which means on average, participant AS scores increased by 0.46 per session.

Chapter 6. Discussion

This thesis had two aims, (i) to investigate the psychometric properties of the AS, and to (ii) investigate how the AS functions as an outcome measure. I addressed the first of these aims in Study 1, assessing the internal reliability and construct validity of the AS, collating data from all publications yielded by the systematic literature search. I also found that no study to date has explored how the AS functions as an outcome measure, nor used it with a clinical sample. Thus, in Study 2 I built on the findings of Study 1, and it was the first to explore the reliability and validity of the AS in a clinical sample.

I addressed the second aim of the thesis, to investigate how the AS functions as an outcome measure, in Study 2, exploring the acceptability and sensitivity to change of the AS. In this chapter, I bring the two studies together and discuss my findings in terms of the two aims of the thesis. I begin by addressing the first aim, discussing the internal reliability and construct validity of the AS. I then move on to the second aim, and discuss the acceptability and sensitivity to change of the AS, as well as the distribution of its scores. I conclude by giving my recommendations for counselling psychologists considering using the AS as an outcome measure in light of my findings, before discussing the implications of this thesis for counselling psychology, and potential directions for future studies.

6.1. The psychometric properties of the AS

6.1.1. Internal reliability

In Study 1, coefficient alphas for AS scores from 30 separate samples, and 8,823 participants, were combined, giving a combined weighted estimate of $\alpha = .84$ (95% CI [.82, .86]).

This is a promisingly high overall estimate according to conventional interpretations (e.g. Chen & Krauss, 2004; Crutzen & Kuntsche, 2013). However, as reliability is a property of the scores of a measure, rather than of the measure itself, coefficient alphas cannot be generalised beyond the sample from which they were calculated. This was clearly illustrated by the very high between-sample heterogeneity in coefficient alphas in Study 1. Despite this, in 28 of the 30 samples, the reported coefficient alphas were between $\alpha = .77$ and $\alpha = .89$, and thus Study 1 did show that for the vast majority of samples in which the internal reliability of the AS has been assessed, coefficient alphas have been high.

In Study 2, AS scores had high internal reliability values: $\alpha = .80$ at baseline, and $\alpha = .88$ at post-therapy (Crutzen & Kuntsche, 2013). These values are similar to the reported values from Study 1, despite the small sample sizes of 61 clients at baseline and 45 at post-therapy. Importantly, this study was the first to report AS coefficient alphas for a clinical sample. Though they cannot be generalised beyond this study, their comparability to the vast majority of the reported coefficient alphas in Study 1 seems to indicate the AS typically has good internal consistency in both clinical and non-clinical samples.

6.1.2. Validity

6.1.2.1. Concurrent validity

In Study 1, I considered every measure the AS has been correlated with in publications yielded by the systematic literature search. As there were too many to report on each one, I focused on the measures I expected to relate to the AS based on the literature, and grouped these by type. I found the AS shared moderate positive correlations with various measures of well-being, operationalising constructs such as life satisfaction, self-esteem, and psychological well-being (e.g. SWLS, Diener *et al.*, 1985; RSE, Rosenberg, 1965; WEMWBS; Tennant *et al.*, 2007).

Similarly, it shared moderate negative correlations with measures of emotional distress, and constructs such as stress, anxiety and depression (e.g. DASS-21, Lovibond & Lovibond, 1995; the perceived stress scale, Cohen & Williamson, 1988; SAS, Zung, 1971). These correlations were all as expected, consistent with the view of humanistic psychology that increased authenticity increases well-being and reduces distress (Rogers, 1959). Thus, I also expected the correlations with the CORE-10 in Study 2 to be of similar magnitude to the measures of distress found in Study 1, and indeed, the AS correlated moderately with the CORE-10 ($r = -.36$) at baseline, sharing 13% variance. This correlation was also important because it is the only one reported from a clinical sample, indicating the relationship between authenticity and distress is similar in both clinical and non-clinical populations.

I also found the correlations between the AS and measures of personality consistent with the design model of the AS (see section 2.4.1.6.). These were all small to moderate, illustrating that the AS is not merely measuring a constellation of personality traits. Finally, the AS also shared moderate to large correlations with other measures of authenticity, further supporting its validity. In all, the correlations between the AS and other measures from both Study 1 and Study 2 are consistent with the conceptual underpinnings of the AS, and support its concurrent validity.

Considering the validity of the AS as an outcome measure, the moderate correlation with the CORE-10 from Study 2 is reassuring. If the correlation had been too large, it would suggest the AS and CORE-10 merely measure opposite ends of a spectrum, with distress on one end and well-being on the other. This would draw into question whether the AS can really add anything new to outcome measurement in counselling psychology. However, the moderate negative correlation indicates the AS and CORE-10 are measuring constructs which, though overlapping, are very different and not merely opposites.

6.1.2.2. Construct validity

I further explored the construct validity of the AS by examining the subscale intercorrelations in both Study 1 and Study 2. The design model of the AS suggests the three subscales, authentic living, accepting external influence, and self-alienation, are three interrelated yet oblique facets of authenticity. This model implies large correlations between each subscale and the AS, and moderate intercorrelations between the subscales. Importantly, these correlations can only support the validity of the AS if each subscale has good internal reliability on its own, indicating that each of the four items in a subscale measure a distinct construct. This was what I found in the meta-analysis, where each subscale correlated highly with the AS ($r = .64 - .82$), and intercorrelated moderately, with correlation coefficients ranging from .34 to .44, and estimated combined coefficient alphas from of $\alpha = .77$, and $\alpha = .82$.

Though Study 2 was conducted with a clinical sample, I expected similar results to those of Study 1, as there was no reason to suspect the design model of the AS would be different in clinical samples. At baseline, each subscale correlated as expected with the AS, and two of the three intercorrelations were as expected. However, the AL and SA correlation was surprisingly non-significant, with a very small coefficient of $r = -.04$, 95% CI [-.29, .21]. The AL and SA being unrelated at baseline diverges markedly from both the theoretical design model of the AS, and what was found in the meta-analysis, where their estimated combined correlation coefficient was $r = -.36$, 95% CI [-.41, -.32]. I investigated whether this lack of a correlation could be due to low statistical power, considering the sample size was 61. Thus, as advocated by O’Keefe (2007), I calculated the statistical power of the correlation for a sample size of 61, assuming the population effect was the same as the combined correlation coefficient from Study 1, $r = -.36$. This gave an estimated power of .89. Thus, assuming the AL and SA are indeed related in accordance with the design model of the AS, the analysis should have had sufficient power to detect the effect.

This finding could indicate that the entire factor structure of the AS is problematic. The design model of the AS implies the three factors should be distinct, yet correlated. If one is orthogonal to another, it questions whether they can be seen as parts of a whole, and thus whether they can coherently be added together to get a full-scale score, casting doubt over the validity of the AS. However, at post-therapy, all correlations were as expected, with the AL and SA correlating significantly at $r = -.44$ (95% CI [-.65, -.17]).

One potential explanation of the interesting lack of a correlation between AL and SA at baseline, is that it is a reflection of it being a clinical sample, indicating that for people who are more distressed, the extent to which they are alienated from themselves and are living authentically are unrelated. In other words, someone who is psychologically distressed might be highly self-alienated, yet still allow what they are feeling and thinking to be reflected in their words and behaviour. If this is the case, then the moderate correlation at post-therapy might reflect that psychotherapy somehow reconnects these two aspects of authenticity.

Another point to consider is whether this finding points back to the difference between state and trait authenticity, and might indicate that not all the subscales are best understood as traits. It may be argued that whereas authentic living (AL) appears reflective of a stable trait, self-alienation (SA) might better understood as a contextual state. This might account for the lack of a correlation at baseline, as distress might cause disruption in the inner experience of authenticity at a day-to-day level.

It is also worth noting that despite the seemingly high power of this analysis, it cannot be ruled out that this surprising finding was due to random error this study is vulnerable to due to its small sample size (Lavrakas, 2008). Regardless, more research is needed to further explore the structure of the as, and how the AL and SA subscales relate, in larger clinical samples.

6.1.2.2. Factorial validity

Finally, I explored the construct validity of the AS in Study 2 by looking at its factor structure using CFA. The CFA found that the three-factor hierarchical model of the AS, did not fit the baseline data, the opposite of the CFA findings reported by Wood *et al.* (2008), Grégoire *et al.* (2014), and Grijak (2017). Though the fit indices appeared reasonable, the standardised regression weight between the AS and the AEI, which can only range from -1 to 1, was -2.00. According to Kline (2016), this is a Heywood case, and indicates that the Amos software (Arbuckle, 2017) could not calculate a useable solution. According to Kline (2016), Heywood cases can be due to small sample sizes ($N < 100$), which, considering the sample size in this study was 61, is a potential source of the issue.

However, considering the AL and SA were unrelated at baseline, it is not surprising the CFA failed to support the tripartite model of the AS. Similarly, though there was insufficient data to run a CFA ($n = 45$) with post-therapy scores, it is likely it would have found the tripartite model a good fit for the data had the sample been larger, as the AL to SA correlation was as expected post-therapy. Regardless, the factor structure of the AS needs further researching in larger clinical samples, to support its validity. If the lack of fit persists in other samples, it would indicate the design model of the AS is not appropriate, drawing into question whether it is reasonable to consider the three subscales facets of a whole.

6.1.3. Summary of the psychometric properties of the AS from Studies 1 and 2

Taken together, Studies 1 and 2 both found AS scores had excellent internal reliability in both clinical and non-clinical samples. However, though there is nothing in these findings that might suggest this will not also be the case for future studies, it is still important that reliability is assessed for all samples, illustrated by the very high between-study heterogeneity in Study 1.

In terms of construct validity, Study 1 found the AS correlated in expected ways with a number of measures of well-being, psychological distress, personality and authenticity, and Study 2 found the correlation with the CORE-10 was also as expected based on the design model of the AS. The subscale intercorrelations in Study 1 further supported the construct validity of the AS, but the same correlations in Study 2 did not, as the AL and SA were unrelated at baseline. Similarly, the CFA failed to support the three-factor model of the AS. More research is needed to explore this relationship in larger clinical samples to determine whether these two problematic findings are due to the small sample size, or that the structure of the AS is actually different in clinical samples. If these findings are replicated in other clinical samples, it questions the validity of the AS.

In all however, studies 1 and 2 indicate the AS has excellent psychometric properties, sufficiently so to be used as an outcome measure.

6.2. The AS as an outcome measure

The second aim of this thesis, to investigate how the AS functions as an outcome measure for psychotherapy, was the novel contribution of this thesis to counselling psychology. According to the systematic review, Study 2 is the first to use the AS in such a capacity. Below I discuss the psychometric properties most relevant to its use as an outcome measure: its acceptability, and pivotally, its sensitivity to change, as well as its score distributions.

6.2.1. Acceptability

Arguably the first point of interest in considering an outcome measure is its acceptability, whether clients actually complete it, and what they think of it. The systematic review in Study 1

revealed that no publications to date have reported on the acceptability of the AS, and thus this study is the first to do so.

I assessed acceptability in Study 2, by looking at how many measures, out of the total collected, had omitted items. For the AS, 159 measures were collected in all, and four of those had one omitted item. Two of these had omitted item 2: 'I don't know how I really feel inside', whereas the other two had omitted item 3: 'I am strongly influenced by the opinions of others' and item 4: 'I usually do what other people tell me to do' (Wood *et al.*, 2008) These four omitted items gave an overall a completion rate of 99.97%, suggesting clients in general have no issues with completing the AS.

Completion rates, however, do not say much about the experience of completing a measure, which one would typically need qualitative methods to explore, conducting interviews with clients. The challenge with that approach is that it is a lengthy process, and it might be difficult to get clients to agree to interviews. The TQHS (Di Malta *et al.*, unpublished) allowed me ask about clients' experiences of completing the AS in self-report form, and use quantitative methods to analyse them. I found that the AS scored right in the middle of the TQHS scoring scale, with a POMP score of 52 %. Thus, it appears clients on average neither liked nor disliked the AS. Moreover, the TQHS POMP score for the AS was comparable to the 53% score of the CORE-10, indicating the acceptability of the two measures is similar. Coupled with the high completion rates, these TQHS results indicate the AS is indeed an acceptable measure.

6.2.2. Response distribution

I also assessed the response distribution of the AS scores, a concept closely linked to sensitivity to change, as the presence of floor or ceiling effects can limit the sensitivity of the measure (Lipsey, 1990; Hessling *et al.*, 2004). No such effect was detected when considering the

POMP scores of the AS and its subscales. The mean POMP score for the AS at baseline was 50%, right in the middle of the scoring scale, which, coupled with the POMP *SD* of 15.4, suggests clients used the entire scoring scale of the AS, on either side of the mean. Equally, no floor or ceiling effects were present in the POMP scores of the subscales, though the AL did have a higher POMP than the other two, likely due to it being a positively worded scale.

Looking at the response distribution of the individual items, none had any clear floor or ceiling effects, though three of the items on the AL subscale (items 8, 9 and 11) had no clients use the lowest response category. This might be an effect of socially desirable responding, as ticking '1' on item 8 meant saying 'I always stand by what I believe in' 'does not describe me at all', with items 9 and 11 similar to item 8. Two of the items on the AL subscale also tended towards ceiling effects, with 24% of participants picking the highest response category for item 1, and 25% for item 8. Similarly, 25% of participants scored the lowest category on item 10. However, these effects should be unproblematic, particularly as the overall measure showed no such effects.

6.2.3. Sensitivity to change

Sensitivity to change was the pivotal point of this research and is vital for any outcome measure. According to Study 1, coefficient alpha values have been reported for the AS in 30 different samples, and 26 studies have reported correlations between the AS and other measures. Its sensitivity to change however, was unknown prior to this research. I explored the sensitivity of the AS in two ways, using both a simple t-test to compare baseline and post-therapy scores, as recommended by Stratford and Riddle (2005), and by using MLM, as advocated by Tasca and Gallop (2009). The problem with the first method was that I had to limit the data to only baseline and post-therapy scores, thus ignoring 33% of the data recorded throughout therapy, and limiting the analysis to 41 out of 67 clients. Using MLM however, allowed me to assess sensitivity moving

beyond just baseline and post-therapy comparison, and include all data for all clients in the analysis, regardless of how many measurements each had completed.

The t-test method established that the AS is sensitive to change, showing that AS scores increased significantly from baseline to post-therapy. This is a moderate change according to suggested benchmarks (Husted *et al.*, 2000), with an SRM change score of 0.62. Pivotaly, the MLM echoed these findings, showing that clients' AS scores improved by 0.46 per session on average.

6.2.3.1. Clinically significant (CSC) and reliable change (RC)

Of the 41 clients with change data, only 29 % showed reliable improvement on the AS, indicating that for 71% of clients, the change in their AS scores could be due to measurement error rather than actual experienced change in authenticity. Similarly to RC, 24% of the 41 client who had both baseline and post-therapy scores showed CS improvement on the AS, which means 10 clients had reliably moved from AS scores typical of a clinical population at baseline, to scores typical of a non-clinical population post-therapy. Two other clients had moved from the clinical score category to the non-clinical, but this change was not reliable. One client (2 %) moved from the non-clinical to the clinical score category, but again this was not reliable deterioration, indicating the client did not actually deteriorate by much, but rather started right above the cutoff point for CSC, and ended just below it.

To compare the AS with the CORE-10 in terms of the proportion of clients who showed CSC and RC, I examined these properties in the 19 clients who had both AS and CORE-10 change scores. In this sample, 37% of clients showed reliable improvement on the AS, whereas 47% showed it on the CORE-10. Interestingly, though the proportion of clients showing this on the CORE-10 is higher than on the AS, it is not as high as in previous publications reporting on RC

for the CORE-10 and the CORE-OM (Barkham *et al.*, 2013; Evans *et al.*, 2002). In fact, Evans *et al.* (2002) reported that 74% of clients in their primary care counselling sample ($n = 125$) showed RC on the CORE-OM. This seems to indicate that the sample in Study 2 did not show as much change in general as the sample used by Evans *et al.* (2002), and that though only 37% showed reliable improvement on the AS, this would have been substantially higher in their sample.

In terms of CSC, 32% showed CS improvement on the AS, versus 45% on the CORE-10, which is similar to the counselling sample in Evans *et al.* (2002). No clients showed CS deterioration on either measure.

It is worth noting that the small sample size in this study may have affected both RC and CSC proportions. This is because both RCrit and the CSC criterion are functions of the *SD* of the baseline mean scores, and as the *SD* is inversely related to sample size, the low percentage of clients experiencing RC and CSC may be due to the small sample sizes in this study.

6.2.3.2. Comparing the sensitivity of the AS with that of the CORE-10

From the above section, it is clear a greater proportion of clients showed reliable and clinically significant change on the CORE-10 than on the AS in this sample ($n = 19$). This might reflect that authenticity, as measured on the AS, is a more stable construct which takes relatively longer to change than symptoms of psychological distress measured on the CORE-10. Indeed, the wording of the two scales differs, with the CORE-10 asking clients how they have felt the past week, whereas the AS does not give a timeframe to consider. However, it may also be interpreted as the AS simply not being as sensitive to change as the CORE-10.

That being said, the differences are not as pronounced as they might seem. Four of the five clients who showed CSC on the CORE-10, but not on the AS, started above the cutoff, and could thus not show CS improvement, and the final client showed reliable improvement, but had a post-

therapy score just one below the cutoff. Moreover, when comparing baseline and post-therapy scores in the same 19, the SRM effect sizes of the change were similar: the AS had an SRM effect size of 0.73, whilst the CORE-10 had an SRM of 0.77. Moreover, the standardised change on these two measures were not significantly different.

Taken together, it seems reasonable to conclude that the AS and CORE-10 in this sample are sensitive to change at a comparable level. Since the CORE-10 is already an established and widely used outcome measure (CORE users, n.d.), this is an important point, and one that speaks in favour of the AS also being employed as an outcome measure. However, an important caveat is that these analyses were conducted in a very small sample ($n = 19$), one too small to provide conclusive conclusions, and further research is needed to explore these issues.

6.2.3.3. Change scores and construct validity

To further assess the construct validity of the AS, I explored how change on the AS related to change on the CORE-10. Based on the literature, discussed extensively in chapter 1 and the narrative synthesis in Study 1, holding authenticity and psychological distress to be inversely, but obliquely related, I hypothesised a moderate negative correlation. In Study 2, the correlation between the change scores was $r = -.35$ (95% CI [-.69, .12]), but only nearing significance ($p = .070$). However, a power analysis found that a sample size of 49 would be needed to detect an effect of this magnitude with .80 power, indicating that low power due to the small sample size was the likely cause of the non-significant result.

I also assessed construct validity by investigating what proportion of change on the AS could be attributed to each of the subscales. Interestingly, post-therapy AL scores were no higher than at baseline, whereas AEI scores had increased by 9.4% and SA scores by 19.4%. AL had an SRM effect size of 0.03, whilst AEI had an SRM of 0.50, and SA of 0.69. It was not clear from

the conceptual design model of the AS what to expected from this analysis, and thus I had no clear hypothesis. However, these findings do seem congruent with Rogerian theory, and thus the design model of the AS, as Rogers (1959) did not see change happening simultaneously on these three facets of authenticity, but sequentially. Clients were thought to become less self-alienated (SA) through therapy, which had a causal effect on the extent to which they expressed their thoughts and feelings (AL), as well as their locus of evaluation, which is the construct the AEI is largely based on (Barrett-Lennard, 1998; Wood *et al.*, 2008). This finding then, could reflect that change on the SA subscale happens initially, and has a causal effect on AL and AEI change. If this is the case, then the lack of any change on the AL might simply be because it happens later, and a follow-up measurement a while after the conclusion of therapy would be needed to register it.

It is also worth considering if the issue of whether all the subscales are indeed best understood as traits can account for these findings. If SA is more reflective of state authenticity, and AL is more reflective of trait authenticity, then it would seem to make sense that change on SA would be considerably greater than on AL, as states are less stable than traits.

6.2.4. Summary of how the AS functions as an outcome measure

In Study 2, I addressed the second aim of this thesis, assessing how the AS functions as an outcome measure. I found that it was acceptable, with very high completion rates and good TQHS scores comparable to the CORE-10, indicating clients were fairly neutral about completing it. In assessing score distributions, I found the AS was not affected by floor or ceiling effects, and clients used the entire scoring scale for the majority of items. Pivotaly, I also found the AS is sufficiently sensitive to register change clients experience over the course of therapy, with scores significantly higher after therapy than before, and clients' AS scores increasing by 0.46 per session on average.

It can be concluded that the sensitivity of the AS is similar to that of the CORE-10, despite more clients showing reliable and clinically significant change on the CORE-10.

6.3. Implications for counselling psychology

For counselling psychologists, or any other clinicians, considering the AS as an outcome measure, this thesis indicates it is sufficiently psychometrically sound, acceptable to clients, and sensitive to change, to be used in such a capacity. Considering its conceptual basis in humanistic psychology (Wood *et al.*, 2008), it may be of particular interest to clinicians practicing within these models, along with other perspectives that hold authenticity as a key outcome of psychotherapy (e.g. Horney, 1951; Kohut, 1971; May, 1981; Perls, 1969; Winnicott, 1965).

The AS could also be a useful outcome measure for practitioners and services wishing to use a well-being focused measure, consistent with positive psychology, instead of or in addition to a measure of psychological distress, such as the CORE-10. Indeed, though the AS and CORE-10 were moderately related, their change scores were largely divergent, suggesting the AS could add valuable breadth to outcome measurement. However, it is important to bear in mind the AS does not assess risk, and might therefore be more suited to be used in addition to a measure that assesses risk. Furthermore, the constructs the AS aims to measure, as well as the wording of the items, suggests it does not need to be a sessional measure, but can be used at intervals, as I have done in this research (Wood *et al.*, 2008).

The AS also seems an ideal measure for researchers interested in exploring how authenticity relates to other constructs, or the psychotherapeutic and aetiological theories of humanistic psychology (Rogers, 1959).

Though more research is needed in assessing the AS, it could be useful to health research at a governmental level, considering the importance key counselling psychology perspectives have

placed on authenticity, along with the strong relationship between the AS and measures of well-being and psychological distress found in Study 1. It could, for example, add an important mental well-being and positive psychology focus to the IAPT national data set (NHS Digital, 2019), and supplement the WEMWBS in the National Health Survey (NatCen for Social Research, 2018). Moreover, as mental health issues among children and young people is a growing issue in the UK (e.g. Pitchforth, Fahy, Ford, Wolpert, Viner, & Hargreaves, 2018), the AS might be useful in exploring and addressing these issues, as the aforementioned counselling psychology perspectives see it as pivotal to well-being, and in understanding psychological distress and dysfunction (Rogers, 1959; Wood *et al.*, 2008). Research with the AS may thus prove relevant to educational and health policy, for example in teaching children and young people the importance of authenticity, helping them to be less alienated from what is happening in them at a deep level and more able to express their thoughts and feelings to the world around them.

6.4. Limitations

6.4.1. Sample size

The major limitation throughout Study 2, which is important to be aware of when interpreting its findings, is its small sample size. Though 67 participants were recruited overall, for most analyses the number was considerably smaller. This was particularly true for analyses involving the CORE-10, as not all sites were able to include it in their protocols. This was a natural constraint of this being a practitioner-doctoral thesis, and therefore greatly limited in its scope. This meant all services were involved on a volunteer basis, offering of their time, energy and resources to contribute. In addition, the nature of doing research with clients receiving psychotherapy is that there is a high degree of drop-out compared to other research designs. Indeed,

27 % of clients who completed the AS, and 19 % of those who completed the CORE-10, only completed one measurement.

The danger of a small sample size is that it limits statistical power and leaves the study more vulnerable to random type II error, that is, failing to find a real effect that is present (Acheson, 2010; Lavrakas, 2008). Thus, it limits the conclusions that can be drawn from this study, and it is best seen as a pilot study, laying the foundation for future and more comprehensive research.

See section 2.4.2.4. for limitations of Study 1.

6.5. Directions for future research

As this is the first study to have used the AS with clinical samples, and the first to have started to explore it as an outcome measure for psychotherapy, it has opened up several directions for future research. These include further exploration of how the AS functions as an outcome measure, and exploring claims made by key counselling psychology perspectives (e.g. Rogers, 1959, Horney, 1951; May, 1981) regarding how authenticity relates to distress and how it is affected by psychotherapy.

6.5.1. The present research as a pilot study

The present research was designed to be a pilot study, giving an indication of how the AS functions as an outcome measure in a small, clinical sample. The natural next step then would be to conduct this research again with a similar design, but with a larger dataset, which would allow for more confidence in the findings and the subsequent conclusions. This has indeed been the plan from the outset of this research, and the services involved have continued to collect data with this aim.

6.5.2. Exploring how psychotherapy impacts authenticity

As has been discussed, several key perspectives within counselling psychology hold that the aim of psychotherapy is increased authenticity (e.g. Horney, 1951; May, 1981; Kohut, 1971; Rogers, 1959; Winnicott, 1965; Yalom, 1980). The conceptual model of the AS and its factor structure allows us to explore how psychotherapy impacts each of its three subscales. In Study 2 I found that change on the SA was significantly greater than on the other two subscales, which might be due to SA change happening first and having a causal effect on AL and AEI, which is consistent with humanistic psychological theory (Barrett-Lennard, 1998; Rogers, 1959). It would be very interesting for future research to look more in-depth at how psychotherapy impacts each of these three constructs, and whether change on the SA does indeed have a causal effect on AL and AEI change. This might involve administering the AS a few months after the conclusion of therapy, and seeing if the AL and AEI change scores have caught up with those of the SA.

6.5.3. Exploring the relationship between authenticity and distress

Closely linked to how psychotherapy affects authenticity, is how authenticity relates to psychological distress. From the aforementioned perspectives within counselling psychology, the link between the two constructs appears to be causal: when an individual becomes more authentic, this reduces elements of distress such as anxiety and depression (Rogers, 1959). This research did find that distress and authenticity, operationalised by the CORE-10 and AS, were linked, correlating moderately at baseline and sharing 13% variance. However, this was merely cross-sectional, and the design and scope of the project meant it was not possible to probe any causal links between the two measures. Boyraz *et al.* (2014) on the other hand, did find a unidirectional causal link between AS scores at baseline and DASS-21 scores (Lovibond & Lovibond, 1995) six weeks later, providing support for the hypothesised causal link between authenticity and distress.

However, their study did not involve any intervention: the measures were simply administered at baseline and at six weeks. Thus, beyond this thesis, no study has investigated how psychotherapy affects the relationship between these two variables.

6.5.4. The AS, CORE-10, and relapse

Another related direction for future research would be looking at how lasting the effects of psychotherapy are, and how well the AS predicts this as compared to a distress or symptom focused measure. From counselling psychology perspectives (e.g. Horney, 1951; May, 1981; Rogers, 1959; Winnicott, 1965), one might hypothesise that change measured on the AS is relatively stable and lasting, and potentially more so than on symptom focused measures. It would have been very interesting to see whether clients who experienced change on the AS experienced more lasting change than those who only experienced change on the CORE-10, and equally whether change on both measures predicts better lasting outcome than change on one measure.

Chapter 7. Conclusions

7.1. Summary

This thesis sought to meet two primary aims, (i) investigate the psychometric properties of the AS, and (ii) investigate how the AS functions as an outcome measure. I addressed these aims in two studies. Study 1 was a systematic review of the literature, culminating in a narrative synthesis, collating the evidence regarding the concurrent validity of the AS, and a meta-analysis, collating the coefficient alphas, means scores and subscale intercorrelation coefficients from all relevant studies. The narrative synthesis identified 26 publications that have reported correlations between the AS and measures of well-being, distress, personality and authenticity, finding them to provide support for the construct validity of the AS. The meta-analysis found that for the vast majority of publications, the AS has excellent internal reliability, with an overall estimated coefficient alpha of $\alpha = .84$. Furthermore, the estimated subscale intercorrelations provided support for the factor structure reported by Wood *et al.* (2008), and POMP scores indicated there were no problematic floor and ceiling effects. AS expected, there was a large degree of variance between studies, which means the findings of the meta-analysis cannot be directly generalised to further populations, but it is reasonable to conclude Study 1 indicates the AS is sufficiently psychometrically sound to function as an outcome measure.

The second study aimed to directly explore how the AS functions as an outcome measure for counselling psychology, assessing its acceptability and sensitivity to change, as well as its construct validity. It had good acceptability, with near perfect completion rates and TQHS scores comparable to the CORE-10, close to the middle of the scale, and there appeared to be no problematic floor or ceiling effects. Its internal reliability was also sufficient at both baseline ($\alpha =$

.80) and post-therapy ($\alpha = .88$), and correlations with the CORE-10, as well as its subscale intercorrelations, supported its validity. Finally, and pivotally, the AS was sensitive to change, with clients' scores significantly higher post-therapy than at baseline, and MLM finding clients improved by 0.46 points per session on average. Moreover, its sensitivity was found to be comparable to that of the CORE-10.

However, a major limitation with Study 2 was its small sample size, limiting the conclusions that can be drawn from its findings. This may be the reason for the two problematic findings, with the AL and SA subscales not correlating at baseline, and the CFA not finding a solution, results that question the validity of the AS. Thus, though this study has opened up several interesting directions for future research, an important one is to assess the AS as an outcome measure in larger clinical samples, providing researchers and practitioners who are considering using it in some capacity, further confidence in its properties.

7.2. Reflexivity

The process of conducting this research and writing this thesis, and all the different elements that have gone with it, has had a great impact on me, and been one of the most significant processes of my life. Coming to the end of it now, it is not easy to reflect on and express just how I have developed and what I have learnt. The most significant learning has certainly been personal, persevering through moments of self-doubt and feeling overwhelmed; liaising with therapists, services and academics whilst feeling under-qualified and insecure. Pivotaly, I learnt to balance zooming out to get a meta-perspective, and zooming in so I could simply focus on the next step, rather than the mountain I have yet to climb. Beyond that, this project has of course greatly broadened my conceptual, methodological and statistical understanding of counselling psychology and the research process.

The perhaps greatest conceptual learning has come from reflecting on, reading about, and finally formulating, my epistemological position. This taught me the importance of taking a meta-perspective in research, of examining the assumptions that underlie every research decision, and every word I write. The epistemology was undoubtedly the most challenging part of the writing process, particularly as it is an element of this research that has changed considerably over the course of the project. At first, I saw epistemology simply as a checklist item to be ticked off, without any real impact on the research process. Now however, I believe it to be the most fundamental starting point for any research. When I began this project three years ago, I held to a positivist view, though much of it was implicit assumption. The notion of measuring authenticity, or any mental attributes, seemed little more problematic than measuring the temperature, and I believed that as long as I used quantitative methods, my findings would be objective and impartial. The research process, and the significant amount of reading that has gone with it, has problematised this position for me. I had to consider what authenticity is, whether it has reality, and whether it exists separately from its socially observable manifestations. I also started considering whether it can be measured usefully, that is, can a person's experiences be meaningfully reduced to 12 numbers on a questionnaire? These questions eventually lead me to move from a positivist view, to a realism-pragmatism view of psychological measurement.

All in all, this process and the past three years has left me feeling grateful to the people who have enabled me to complete it, my family and my supervisory team. It has taught me so much, and left me with a desire to conduct further research and contribute to the field of counselling psychology.

References

- Acheson, A. (2010). Sample size. In N. J. Salkind (Ed.) *Encyclopedia of research design* (pp. 1300-1301). Thousand Oaks, CA: Sage Publications Inc.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267-281). Budapest: Akademiai Kiado.
- Akın, A. & Dönmezogullari, C. (2010) The validity and reliability of Turkish version of the Authenticity Scale. Paper presented at the 2nd International Congress of Educational Research, 29 April–2 May, Antalya, Turkey.
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- APA Presidential Task FORCE on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, 61(4), 271-285.
- Apostolou, A., Ward, A., Yakeley, J. (2016). Outcome Measures for Psychodynamic Psychotherapy Services. Royal College of Psychiatrists.
- Arbuckle, J. L. (2017). Amos, Version 25.0 [Computer software]. Chicago: IBM SPSS.
- Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto alexithymia scale: I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, 38, 23–32.
- Baker, T. L., Rapp, A., Meyer, T., & Mullins, R. (2014). The role of brand communications on front line service employee beliefs, behaviors, and performance. *Journal of the Academy of Marketing Science*, 42(6), 642-657.

- Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J. A., & Nielsen, S. L. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose–effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology, 77*(2), 203.
- Barkham, M., Bewick, B., Mullin, T., Gilbody, S., Connell, J., Cahill, J., Mellor-Clark, J., Richards, D., Unsworth, G., & Evans, C. (2013). The CORE-10: A short measure of psychological distress for routine use in the psychological therapies. *Counselling and Psychotherapy Research, 13*(1), 3-13.
- Barkham, M., Evans, C., Margison, F., McGrath, G., Mellor-Clark, J., Milne, D., & Connell, J. (1998). The rationale for developing and implementing core batteries in service settings and psychotherapy outcome research. *Journal of Mental Health, 7*, 35-47.
- Barkham, M., Hardy, G.E., & Mellor-Clark, J. (Eds.) (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies*. Chichester: Wiley.
- Barnette, J.J. (1996). Responses that may indicate nonattending behaviors in three self-administered educational surveys. *Research in the Schools, 3*(2), 49–59.
- Barrett-Lennard, G. T. (1998). *Carl Rogers' helping system: Journey & substance*. London: Sage.
- Beck A.T. (1970). Cognitive therapy: Nature and relation to behavior therapy. *Behavior Therapy, 1*, 184–200.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571.
- Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*(4), 1088-1101.
- Bhaskar, R. (1997). On the ontological status of ideas. *Journal for the Theory of Social Behaviour, 27*(2-3), 139-147.

- Bickman, L., Rosof-Williams, J., Salzer, M. S., Summerfelt, W. T., Noser, K., Wilson, S. J., & Karver, M. S. (2000). What information do clinicians value for monitoring adolescent client progress and outcomes? *Professional Psychology: Research and Practice*, *31*, 70–74.
- Blaikie, N. (2004). Epistemology. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds) *The SAGE Encyclopedia of Social Science Research Methods* (p. 310). Thousand Oaks, CA: Sage Publications Inc.
- Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *Bmj*, *314*, 572.
- Bohart, A. C. (2013). The actualizing person. In M. Cooper, M. O'Hara, P. F. Schmid, & A. C. Bohart (Eds.), *The handbook of person-centred psychotherapy and counselling* (pp. 84-101). New York, NY, : Palgrave Macmillan.
- Bohart, A. C., O'Hara, M., & Leitner, L. M. (1998). Empirically violated treatments: Disenfranchisement of humanistic and other psychotherapies. *Psychotherapy Research*, *8*, 141–157.
- Bond, M. J., Strauss, N. E., & Wickham, R. E. (2018). Development and validation of the Kernis-Goldman authenticity inventory-short form (KGAI-SF). *Personality and Individual Differences*, *134*, 1-6.
- Bonett, D. G. (2010). Varying coefficient meta-analytic methods for alpha reliability. *Psychological Methods*, *15*(4), 368.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111.
- Boswell, J.F., Kraus, D.R, Miller, S.D., & Lambert, M.J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*, *25*(1), 6-19.

- Boyd, R. (1983). On the current status of the issue of scientific realism. *Erkenntnis* 19, 45-90.
- Boyd, R. (1990). Realism, conventionality, and 'realism about'. In G. Boolos (Ed.) *Meaning and Method* (pp. 171-196). Cambridge: Cambridge University Press.
- Boyratz, G., & Kuhl, M. L. (2015). Self-focused attention, authenticity, and well-being. *Personality and Individual Differences*, 87, 70–75.
- Boyratz, G., Waits, J. B., & Felix, V. A. (2014). Authenticity, Life Satisfaction, and Distress: A Longitudinal Analysis. *Journal of Counseling Psychology*, 61(3), 498–505.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3), 185-216.
- British Psychology Society, (2017). *BPS practice guidelines* (3rd edn.) [Online]. Available from [https://www.bps.org.uk/sites/bps.org.uk/files/Policy/Policy%20-%20Files/BPS%20Practice%20Guidelines%20\(Third%20Edition\).pdf](https://www.bps.org.uk/sites/bps.org.uk/files/Policy/Policy%20-%20Files/BPS%20Practice%20Guidelines%20(Third%20Edition).pdf) [Accessed: 25 February 2019].
- Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult attachment. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 46–76). New York: Guilford Press.
- Brislin, R.W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3), 185-216.
- Brown, K. W. & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, 84, 922–948.
- Cai, H., Sedikides, C., & Jiang, L. (2013). Familial self as a potent source of affirmation: Evidence from China. *Social Psychological and Personality Science*, 4(5), 529-537.

- Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavalley, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology, 70*(1), 141–156.
- Carlier, I. V., Meuldijk, D., Van Vliet, I. M., Van Fenema, E., Van der Wee, N. J., & Zitman, F. G. (2012). Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *Journal of evaluation in clinical practice, 18*(1), 104-110.
- Carlier, I. V. & van Eeden, W. A. (2017). Routine Outcome Monitoring in Mental Health Care and Particularly in Addiction Treatment: Evidence-Based Clinical and Research Recommendations. *Journal of Addiction Research and Therapy, 8*(4), 1-7.
- Carver, C. S. & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology, 67*, 319–333.
- Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour research and therapy, 47*(11), 910-920.
- Clement, P. W. (1994). Quantitative evaluation of 26 years of private practice. *Professional Psychology: Research and Practice, 25*(2), 173-176.
- Clift, S., Manship, S., & Stephens, L. (2017). Further evidence that singing fosters mental health and wellbeing: the West Kent and Medway project. *Mental Health and Social Inclusion, 21*(1), 53-62.
- Cloninger, C., Svrakic, D. M., & Przybeck, T. R. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry, 50*, 975–990.

- Chen, P. Y. & Krauss, A. D. (2004). Cronbach's alpha. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds) *The SAGE Encyclopedia of Social Science Research Methods* (pp. 586-587). Thousand Oaks, CA: Sage Publications Inc.
- Cherek, D. R. (1981). Effects of smoking different doses of nicotine on human aggressive behavior. *Psychopharmacology*, 75, 339–345.
- Cicero (1923). Laelius on Friendship. In *Laelius De Amicitia*. (pp. 109-211). Boston: Harvard University Press and Loeb Classical Library. Available: http://penelope.uchicago.edu/Thayer/E/Roman/Texts/Cicero/Laelius_de_Amicitia/text*.html
- Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10, 101-129.
- Cohen J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Cohen, J. (1988). The effect size index: D. In J. Cohen (Ed.), *Statistical power analysis for the behavioral sciences* (Vol. 2, pp. 284–288). Mahwah, NJ: Erlbaum.
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate behavioral research*, 34(3), 315-346.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385–396.
- Cohen, S., & Williamson, G. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan & S. Oskamp (Eds.), *The social psychology of health* (pp. 31–67). Newbury Park, CA: Sage.

- Comte, A. (1974). Aim of the course: General considerations on the nature and importance of positive philosophy. In S. Andreski (Ed.) *The essential Comte* (pp. 19–41). London: Croom Helm.
- Cooper, M. (2008). *Essential Research Findings in Counselling and Psychotherapy. The Facts are Friendly*. London: Sage Publishing.
- Cooper, M. (2013). Developmental and personality theory. In M. Cooper, M. O’Hara, P.F. Schmid, & A.C. Bohart (Eds.), *The handbook of person-centred psychotherapy and counselling* (2nd edn.). (pp. 118-135). London: Palgrave Macmillan.
- CORE Users. (n.d.). Retrieved October 8, 2018, from http://www.coreims.co.uk/About_CORE_Users.html
- Corrie, S. (2010). What is evidence. *Handbook of counselling psychology*, 44-61.
- Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, Florida: Psychological Assessment Resources.
- Cramer, D., & Howitt, D. L. (2004). *The Sage dictionary of statistics: a practical resource for students in the social sciences*. London: Sage.
- 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.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Cruickshank, J. (2012). Positioning positivism, critical realism and social constructionism in the health sciences: A philosophical orientation. *Nursing Inquiry*, 19(1), 71-82.

- Crutzen, R. & Kuntsche, E. (2013). Validation of the four-dimensional structure of drinking motives among adults. *European Addiction Research, 19*, 222–226.
- Crutzen, R., & Peters, G. J. Y. (2017). Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review, 11*(3), 242-247.
- Cumming, G., Fidler, F., Kalinowski, P., & Lai, J. (2012). The statistical recommendations of the American Psychological Association Publication Manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology, 64*, 138-146.
- de Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy Research, 22*, 464–474.
- Dean, K., Walker, Z., & Jenkinson, C. (2018). Data quality, floor and ceiling effects, and test–retest reliability of the Mild Cognitive Impairment Questionnaire. *Patient Related Outcome Measures, 9*, 43-47.
- Deci, E. L. (1980). *The psychology of self-determination*. Lexington, MA: Heath.
- Department of Health (1996) Research and Development. *Towards an evidence-based health service*. London: HMSO.
- Department of Health (1997) *The New NHS: Modern, Dependable*. London: HMSO.
- DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology, 74*, 63–79.
- DeVellis, R. F. (2012). *Scale development: Theory and applications* (Vol. 26). Thousand Oaks: Sage.
- Di Fabio, A. (2014). Authenticity Scale: Un primo contributo alla validazione della versione italiana [Authenticity Scale: A first contribution to validation of the Italian version]. *Counseling: Giornale Italiano di Ricerca e Applicazioni, 7*(2), 231-238.

- Di Fabio, A. & Gori, A. (2016). Developing a new instrument for assessing acceptance of change. *Frontiers in Psychology, 7*, 1-10.
- Diener, E., Emmons, R.A., Larsen, R.J., & Griffin, S. (1985). The Satisfaction with Life Scale. *J Personality Assessment, 49(1)*, 71-75.
- Di Malta, G. S., Pauli, G., & Cooper, M. (2018). The Therapy Questionnaire Helpfulness Survey (TQHS). Unpublished manuscript.
- Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of economic psychology, 29(1)*, 94-122.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj 315*, 629-634.
- Eggers, F., O'Dwyer, M., Kraus, S., Vallaster, C., & Guldenberg, S. (2013). The impact of brand authenticity on brand trust and SME growth: A CEO perspective. *Journal of World Business, 48(3)*, 340-348.
- Ehlers, A., Maercker, A., & Boos, A. (2000). Posttraumatic stress disorder following political imprisonment: The role of mental defeat, alienation, and perceived permanent change. *Journal of Abnormal Psychology, 109*, 45-55.
- Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry, 33*, 766-771.
- Endler, N. A., Parker, J. D. A., Bagby, R. M., & Cox, B. J. (1991). Multidimensionality and state and trait anxiety: Factor structure of the Endler Multidimensional Anxiety Scales. *Journal of Personality and Social Psychology, 60*, 919-926.

- Erbes, C., Polusny, M. A., Billig, J., Mylan, M., McGuire, K., Isenhardt, C., & Olson, D. (2004). Developing and applying a systematic process for evaluation of clinical outcome assessment instruments. *Psychological Services, 1*(1), 31–39.
- Evans, C., Connell, J., Barkham, M., Margison, F., Mellor-Clark, J., McGrath, G., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry, 180*, 51-60.
- Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health, 1*, 70-72.
- Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J., & McGrath, G. (2000). CORE: Clinical Outcomes in Routine Evaluation. *Journal of Mental Health 9*(3), 247–55.
- Evidence-Based Medicine Working Group. (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *Jama, 268*(17), 2420.
- Eysenck, H. J. (1952). The effects of psychotherapy: an evaluation. *Journal of Consulting Psychology, 16*(5), 319-324.
- Fagg, J., Curtis, S., Stansfeld, S. A., Cattell, V., Tupuola, A. M., & Arephin, M. (2008). Area social fragmentation, social support for individuals and psychosocial health in young adults: Evidence from a national survey in England. *Social science & medicine, 66*(2), 242-254.
- Fat, L. N., Scholes, S., Boniface, S., Mindell, J., & Stewart-Brown, S. (2017). Evaluating and establishing national norms for mental wellbeing using the short Warwick–Edinburgh mental well-being scale (SWEMWBS): findings from the health survey for England. *Quality of Life Research, 26*(5), 1129-1144.

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.
- Fischer, R., & Milfont, T. L. (2010). Standardization in psychological research. *International Journal of Psychological Research*, *3*(1), 88-96.
- Fisher, W. P. (2003). Mathematics, measurement, metaphor and metaphysics II: Accounting for Galileo's "fateful omission". *Theory & Psychology*, *13*, 791–828.
- Fleeson, W., & Wilt, J. (2010). The relevance of Big Five trait content in behavior to subjective authenticity: Do high levels of within- person behavioral variability undermine or enable authenticity achievement? *Journal of Personality*, *78*, 1353–1382.
- Freire, E. (2007). The Strathclyde Inventory: Validation of a person-centred outcome measure. Paper presented at BACP Research Conference, Birmingham, UK.
- Freire, E. S., Elliott, R., & Rodgers, B. (2012). The Strathclyde Inventory-Revised (SI-R): Developing an outcome measure based on Rogers' theory of therapeutic change. Paper presented at 10th Conference for the World Association for Person-Centered and Experiential Psychotherapy and Counseling, Antwerp, Belgium.
- Funder, D. C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science*, *2*, 31–39.
- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, *10*(2), 486-489.
- Gill, J. & Womack, A. J. (2013). The multilevel model framework. In M. A. Scott, J. S. Simonoff, & B. D. Mar(Eds.) *The Sage Handbook of Multilevel Modeling* (pp. 3-18). Thousand Oaks, CA: Sage Publications Inc.

- Gino, F., Norton, M. I., & Ariely, D. (2010). The Counterfeit Self: The Deceptive Costs of Faking It. *Psychological Science, 21*(5), 712–720.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84–96.
- Goldman, B. M., & Kernis, M. H. (2002). The role of authenticity in healthy psychological functioning and subjective well-being. *Annals of the American Psychotherapy Association, 5*, 18-20.
- Goldman, B. M., & Kernis, M. H. (2004). *The development of the authenticity inventory, version 3*. Unpublished data.
- Goldner, L. (2016). Therapists' self-perception, attachment, and relationship: The role of selfobject needs. *Psychoanalytic Psychology, 33*(4), 535-553-
- Gondek, D., Edbrooke-Childs, J., Fink, E., Deighton, J., & Wolpert, M. (2016). Feedback from outcome measures and treatment effectiveness, treatment efficiency, and collaborative practice: A systematic review. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(3), 325-343.
- Govindji, R., & Linley, P. A. (2007). Strengths use, self-concordance and well-being: Implication for strengths coaching and coaching psychologist. *International Coaching Psychology Review, 2*, 143-153.
- Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. *Behaviour Research and Therapy, 8*(3), 249-266.
- Gray, J. A. (1982). Précis of the neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system. *Behavioral and Brain Sciences, 5*, 469–484.

- Grégoire, S., Baron, L., Ménard, J., & Lachance, L. (2014). The Authenticity Scale: Psychometric properties of a French translation and exploration of its relationships with personality and well-being. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, 46(3), 346-355.
- Grijak, Đ. (2017a). Psychometric evaluation of the Authenticity Scale on the sample of students in Serbia. *Psihologija*, 50(1), 85-99.
- Grijak, Đ. (2017b). Authenticity as a predictor of mental health. *Klinička Psihologija*, 10, 23-34.
- Gungor, D., Karasawa, M., Boiger, M., Dincer, D., & Mesquita, B. (2014). Fitting in or Sticking Together: The Prevalence and Adaptivity of Conformity, Relatedness, and Autonomy in Japan and Turkey. *Journal of Cross-Cultural Psychology*, 45(9), 1374–1389.
- Guyon, H., Kop, J-L., Juhel, J., & Falissard, B. (2018). Measurement, ontology, and epistemology: Psychology needs pragmatism-realism. *Theory & Psychology*, 28(2), 149-171.
- Hammersley, M. (2004). Logical positivism. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds) *The SAGE Encyclopedia of Social Science Research Methods* (pp. 586-587). Thousand Oaks, CA: Sage Publications Inc.
- Harter, S. 2002. Authenticity. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology*: 382-394. London: Oxford University Press.
- Harter, S., Marold, D. B., Whitesell, N. R., & Cobbs, G. (1996). A model of the effects of perceived parent and peer support on adolescent false self behavior. *Child Development*, 67, 360–374.
- Hatfield, D.R. & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice*, 35(5), 485-491.

- Hatfield, D.R. & Ogles, B. M. (2007). Why some clinicians use outcome measures and others do not. *Administration and Policy in Mental Health and Mental Health Services Research*, 34, 283-291.
- Health Care and Professions Council (2015). *Standards of proficiency: Practitioner psychologists* (3rd edn). [Online]. Available from: <https://www.hcpc-uk.org/globalassets/resources/standards/standards-of-proficiency---practitioner-psychologists.pdf>. [Accessed: 25 February 2019].
- Heidegger, M. (1996). *Being and Time* (7th edn) (J. Stambaugh, Trans.). Albany, NY: State University of New York Press.
- Hessling, R.M., Traxel, N.M., & Schmidt, T.J. (2004). Ceiling effect. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds) *The SAGE Encyclopedia of Social Science Research Methods* (p. 105). Thousand Oaks, CA: Sage Publications Inc.
- Higgins, J. P. T. & Green, S. (Eds.) (2011). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta- analysis. *Statistics in Medicine*, 21, 1539-1558.
- Hill, C. E., & Corbett, M. M. (1993). A perspective on the history of process and outcome research in counseling psychology. *Journal of Counseling Psychology*, 40(1), 3.
- Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind, intercultural cooperation and its importance for survival* (3rd ed.). New York: McGraw-Hill Professional.

- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Horney, K. (1951). *Neurosis and human growth: The struggle toward self-realization*. London: Routledge.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy. *American Psychologist*, 51, 1059-1064.
- Hoyle, R. H. (2004). Confirmatory factor analysis. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds) *The SAGE Encyclopedia of Social Science Research Methods* (pp. 170-175). Thousand Oaks, CA: Sage Publications Inc.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Huedo-Medina, T. Sanchez-Meca, J. Marin-Martinez, F. & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *CHIP Documents*, 19. URL: http://digitalcommons.uconn.edu/chip_docs/19
- Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: a critical review and recommendations. *Journal of Clinical Epidemiology*, 53, 459-468.
- Hutchinson, D. S. (1995). Ethics. In J. Barnes (Ed.). *The Cambridge companion to Aristotle* (pp. 195–232). Cambridge, UK: The Press Syndicate of the University of Cambridge.
- IBM Corp. Released 2017. IBM SPSS Statistics for Macintosh, Version 25.0 [Computer software]. Armonk, NY: IBM Corp.

- Ionita, G., & Fitzpatrick, M. (2014). Bringing science to clinical practice: A Canadian survey of psychological practice and usage of progress monitoring measures. *Canadian Psychology/Psychologie canadienne*, *55*(3), 187.
- Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, *15*, 336-352
- James, W. (1907). *Pragmatism: A new name for some old ways of thinking*. Cambridge, MA: Harvard University Press.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford.
- Jones, A., & Crandall, R. (1986). Validation of a short index of self-actualization. *Personality and Social Psychology Bulletin*, *12*, 63–73.
- Joseph, S., & Linley, P. A. (2006). Positive psychology versus the medical model. *American Psychologist*, *61*, 332-333.
- Joseph, S. & Murphy, D. (2013). Person-Centered Approach, positive psychology, and relational helping: Building bridges. *Journal of Humanistic Psychology*, *53*(1), 26-51.
- Joseph, S. & Wood, A. (2010). Assessment of positive functioning in clinical psychology: Theoretical and practical issues. *Clinical Psychology Review*, *30*, 830-838.
- Kernis, M. H. & Goldman, B. M. (2006). A multicomponent conceptualization of authenticity: Theory and research. *Advances in Experimental Social Psychology*, *38*, 283-357.
- Kilbourne, A. M., Beck, K., Spaeth-Ruble, B., Ramanuj, P., O'Brien, R. W., Tomoyasu, N., & Pincus, H. A. (2018). Measuring and improving the quality of mental health care: a global perspective. *World psychiatry*, *17*(1), 30-38.

- Kirillova, K., Lehto, X. Y., & Cai, L. (2017). Existential Authenticity and Anxiety as Outcomes: The Tourist in the Experience Economy. *International Journal of Tourism Research*, 19(1), 13–26.
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th ed.). New York: Guildford Press.
- Knaup, C., Koesters, M., Schoefer, D., Becker, T., & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *The British Journal of Psychiatry*, 195(1), 15-22.
- Knoll, M., Meyer, B., Kroemer, N. B., & Schröder-Abé, M. (2015). It takes two to be yourself. *Journal of Individual Differences*, 36, 38-53
- Kohut, H. (1971). *The analysis of the self*. New York, NY: International Universities Press.
- Kovacs, M. (1985). The children's depression inventory (CDI). *Psychopharmacology*, 21, 995–998.
- Kroenke, K., Spitzer, R.L., & Williams, J.B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606-613.
- Kroenke, K., Spitzer, R.L., Williams, J.B., & Löwe, B. (2010). The Patient Health Questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *General Hospital Psychiatry*, 32(4), 345-359.
- Kroenke, K., Wu, J., Yu, Z., Bair, M. J., Kean, J., Stump, T., & Monahan, P. O. (2016). The patient health questionnaire anxiety and depression scale (PHQ-ADS): Initial validation in three clinical trials. *Psychosomatic Medicine*, 78(6), 716.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

- Lambert, M. J. (1983). Introduction to assessment of psychotherapy outcome: Historical perspective and current issues. In M.J. Lambert, E.R. Christensen & S.S. DeJulio (Eds.), *The Assessment of Psychotherapy Outcome*. New York: John Wiley.
- Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M.J. Lambert (Ed.) *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (6th edn., pp.169-218). New York: Wiley.
- Lambert, M. J., Hansen, N. B., Umphress, V. J., Lunnen, K., Okiishi, J., Burlingame, G. M., Huefner, J. C., & Reisinger, C. W. (1996). *Administration and scoring manual for the Outcome Questionnaire (OQ 45.2)*. Wilmington, DE: American Professional Credentialing Services.
- Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, 48(1), 72.
- Lambert, M. J. & Vermeersch, D. A. (2002) Effectiveness of psychotherapy. In M. Hersen, W. Sledge (Eds.), *Encyclopedia of Psychotherapy, Volume 1*. (pp. 709-714). Cambridge, MA, US: Academic Press.
- Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, 55(4), 520.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. 33, 159–174.
- Landy, G., Aref-Adib, G., Kim, H-K., Brooke-Williams, F., Parkes, M., Chapman, J., & McKenna, C. (2013). *The Tyranny of Measurement; Attitudes towards outcome monitoring for psychological therapies*. Available: <https://www.rcpsych.ac.uk/pdf/Aref-Adib%20Golnar%20%20Sept13%20CALC%20CS.pdf>. (05.10.2018).
- Lavrakas, P. J. (2008). Random error. In P. J. Lavrakas (Ed.) *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage Publications Inc.

- Leichsenring, F., & Rabung, S. (2008). Effectiveness of long-term psychodynamic psychotherapy: A meta-analysis. *Jama*, *300*(13), 1551-1565.
- Lenton, A. P., Bruder, M., Slabu, L., & Sedikides, C. (2013). How Does “Being Real” Feel? The Experience of State Authenticity. *Journal of Personality*, *81*(3), 276–289.
- Lenton, A. P., Slabu, L., Sedikides, S., & Power, K. (2013). I feel good, therefore I am real: Testing the causal influence of mood on state authenticity. *Cognition & Emotion*, *27*, 1202-1224.
- Leon, S. C., Martinovich, Z., Lutz, W., & Lyons, J. S. (2005). The effect of therapist experience on psychotherapy outcomes. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, *12*(6), 417-426.
- Levitt, H.M., Stanley, C.M., Frankel, Z., & Raina, K. (2005). An evaluation of outcome measures used in humanistic psychotherapy: Using thermometers to weigh oranges. *The Humanistic Psychologist*, *33*(2), 113-130.
- Liettaer, G. (1993). Authenticity, congruence and transparency. In D. Brazier (Ed.), *Beyond Carl Rogers* (pp. 17-46). London: Constable and Company.
- Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research* (Vol. 19). London: Sage.
- Long, D. M. (2013). Pragmatism, realism, and psychology: Understanding theory selection criteria. *Journal of Contextual Behavioral Science*, *2*, 61-67.
- Lopez, F. G., Ramos, K., Nisenbaum, M., Thind, N., & Ortiz-Rodriguez, T. (2015). Predicting the presence and search for life meaning: Test of an attachment theory-driven model. *Journal of Happiness Studies*, *16*, 103-116.
- Lopez, F. G., & Rice, K. G. (2006). *Preliminary development and validation of a measure of relationship authenticity*. *Journal of Counseling Psychology*, *53*(3), 362–371.

- Luborsky, L. (2001). The meaning of empirically supported treatment research for psychoanalytic and other long-term therapies. *Psychoanalytic Dialogues*, 11(4), 583–604.
- Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33(3), 335-343.
- Lowry, R. (2019) *The confidence interval of rho*. Retrieved from <http://vassarstats.net/rho.html>.
- MacCallum, R.C., Browne, M.W., & Sugawara, H., M. (1996). Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods*, 1(2), 130-49.
- Maltby, J., Wood, A. M., Day, L., & Pinto, D. (2012). The position of authenticity within extant models of personality. *Personality and Individual Differences*, 52(3), 269-273.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, 98(2), 224-253.
- Markus, K. A. & Lin, C. (2010). Construct validity. In N. J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 230-233). Thousand Oaks, CA: Sage Publications Inc.
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, 41, 954–969.
- Marsh, H. W., & Hau, K. T. (1999). Confirmatory Factor Analysis: Strategies for Small Sample Sizes. *Statistical Strategies for Small Sample Research*, 1, 251-284.
- Maslow, A. H. (1968). *Toward a psychology of being* (2nd ed.). New York, NY: Van Nostrand Reinhold.
- Maul, A., Wilson, M., & Irribarra, D. T. (2013). On the conceptual foundations of psychological measurement. *Journal of Physics: Conference Series*, 459, 1-6.
- May, R. (1981). *Freedom and destiny*. New York: Basic Books.

- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*, 175-215.
- McCullough, M. E., Emmons, R. A., & Tsang, J. A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology, 82*, 112–127.
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual for the profile of mood states*. San Diego, CA: Educational and Industrial Testing Services.
- Mendeley Ltd (2018). Mendeley Desktop Version 1.19.2 [Computer software]. Amsterdam, Netherlands: Elsevier.
- Michell, J. (2003). The quantitative imperative: Positivism, naïve realism and the place of qualitative methods in psychology. *Theory & Psychology, 13*(1), 5-31.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine, 6*(7), e1000097.
- NatCen Social Research, University College London. Department of Epidemiology and Public Health. (2018). *Health Survey for England, 2016*. [data collection]. UK Data Service. SN: 8334, <http://doi.org/10.5255/UKDA-SN-8334-1>
- Nezlek, J. B. (2007). A multilevel framework for understanding relationships among traits, states, situations and behaviours. *European Journal of Personality, 21*, 789–810.
- Fat, L. N., Scholes, S., Boniface, S., Mindell, J., & Stewart-Brown, S. (2017). Evaluating and establishing national norms for mental wellbeing using the short Warwick–Edinburgh mental

- well-being scale (SWEMWBS): findings from the health survey for England. *Quality of Life Research*, 26(5), 1129-1144.
- NHS Digital (2016). Improving Access to Psychological Therapies (IAPT): Executive Summary May 2016. Retrieved from:
<https://files.digital.nhs.uk/publicationimport/pub23xxx/pub23389/iapt-month-nov-2016%20-exec-sum.pdf> (27th March, 2018).
- NHS Digital (2019, February 13). Improving access to psychological therapies data set. Retrieved from: <https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/improving-access-to-psychological-therapies-data-set>
- Nunally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Ogles, B. M., Lambert, M. J., & Fields, S. A. (2002). *Essentials of outcome assessment*. John Wiley & Sons Inc.
- O'Keefe, D. J. (2007) Brief Report: Post Hoc Power, Observed Power, A Priori Power, Retrospective Power, Prospective Power, Achieved Power: Sorting Out Appropriate Uses of Statistical Power Analyses. *Communication Methods and Measures*, 1(4), 291-299
- O'Reilly, A., Peiper, N., O'Keefe, L., Illback, R., & Clayton, R. (2016). Performance of the CORE-10 and YP-CORE measures in a sample of youth engaging with a community mental health service. *International Journal of Methods in Psychiatric Research*, 25(4), 324-332.
- Patterson, T. G., & Joseph, S. (2006). Development of a measure of unconditional positive self-regard. *Psychology and Psychotherapy: Theory, Research and Practice*, 79, 557-570.
- Patterson, T. G., & Joseph, S. (2007). Outcome measurement in person-centred practice. In R. Worsley & S. Joseph (Eds.), *Person-centred practice: Case studies in positive psychology* (pp. 200-217). Ross-on-Wye, England: PCCS Books.

- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598–609.
- Payne, G. & Payne, J. (2004). Positivism and realism. In *Key Concepts in Social Research* (pp. 171-174). London: Sage Publications.
- Perls, F. (1969). *Gestalt therapy verbatim*. Boulder, CO: Real People Press.
- Peterson, C. & Seligman, M. E. P (2004). Integrity [Authenticity, Honesty]. In *Character Strengths and Virtues: A Handbook and Classification* (pp. 227-247). Oxford, England: Oxford University Press.
- Phelps, R., Eisman, E. J., & Kohout, J. (1998). Psychological practice and managed care: Results of the CAPP practitioner survey. *Professional Psychology: Research and Practice*, 29, 31–36.
- Piaget, J. (1969). *Judgement and reasoning in the child*. London: Routledge & Kegan Paul.
- Pillow, D. R., Hale Jr, W. J., Crabtree, M. A., & Hinojosa, T. L. (2017). Exploring the relations between self-monitoring, authenticity, and well-being. *Personality and Individual Differences*, 116, 393-398.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2018). nlme: Linear and Nonlinear Mixed Effects Models [Computer software]. R package version 3.1-137. URL: <https://CRAN.R-project.org/package=nlme>.
- Pinto, D. G., Maltby, J., & Wood, A. M. (2011). Exploring the tripartite model of authenticity within Gray's approach and inhibition systems and Cloninger's bio-social model of personality. *Personality and Individual Differences*, 51(2), 194-197.
- Pinto, D. G., Maltby, J., Wood, A. M., & Day, L. (2012). A behavioral test of Horney's linkage between authenticity and aggression: People living authentically are less-likely to respond aggressively in unfair situations. *Personality and Individual Differences*, 52(1), 41–44.

- Pitchforth, J., Fahy, K., Ford, T., Wolpert, M., Viner, R. M., & Hargreaves, D. S. (2018). Mental health and well-being trends among children and young people in the UK, 1995–2014: analysis of repeated cross-sectional national health surveys. *Psychological Medicine*, 1-11.
- Platten, M. J., Newman, E., & Quayle, E. (2013). Self-esteem and its relationship to mental health and quality of life in adults with cystic fibrosis. *Journal of Clinical Psychology in Medical Settings*, 20(3), 392-399.
- Proctor, C., Tweed, R., & Morris, D. (2016). The Rogerian Fully Functioning Person: A Positive Psychology Perspective. *Journal of Humanistic Psychology*, 56(5), 503–529.
- R Core Team (2018). R: A language and environment for statistical computing Version 3.5.2 [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385-401.
- Reis, G. G., Braga, B. M., & Trullen, J. (2017). Workplace authenticity as an attribute of employer attractiveness. *Personnel Review*, 46(8), 1962–1976.
- Reis, G. G., Trullen, J., & Story, J. (2016). Perceived organizational culture and engagement: the mediating role of authenticity. *Journal of Managerial Psychology*, 31(6), 1091-1105.
- Reynolds, C. R., & Richmond, B. O. (1997). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology*, 25, 15–20.
- Riley, R. D., Higgins, J. P., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *Bmj*, 342, 964-967.

- Rise, M. B., Eriksen, L., Grimstad, H., & Steinsbekk, A. (2016). The long-term effect on mental health symptoms and patient activation of using patient feedback scales in mental health outpatient treatment. A randomised controlled trial. *Patient Education and Counseling*, 99(1), 164-168.
- Robinson, O. C., Lopez, F. G., Ramos, K., & Nartova-Bochaver, S. (2013). Authenticity, Social Context, and Well-Being in the United States, England, and Russia: A Three Country Comparative Analysis. *Journal of Cross-Cultural Psychology*, 44(5), 719–737.
- Rodriguez, M. C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods*, 11(3), 306.
- Rogers, C. R. (1951). *Client-centered psychotherapy*. Boston: Houghton-Mifflin.
- Rogers, C.R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, 21, 95-103.
- Rogers, C.R. (1959). A theory of therapy, personality and interpersonal relationships as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: A study of science* (pp. 184-256). New York: McGraw-Hill.
- Rogers, C.R. (1961a). *On Becoming a Person: A Therapist's View of Psychotherapy*. Boston, MA: Houghton Mifflin.
- Rogers, C.R. (1961b). The process-equation of psychotherapy. *American Journal of Psychotherapy*, 15, 27-45.
- Rogers, C. R. (1963). Actualizing tendency in relation to "Motives" and to consciousness. In M. R. Jones (Ed.), *Nebraska symposium on motivation* (pp. 1-24). Oxford, England: U. Nebraska Press.
- Rogers, C. R. (1966). Client-centered therapy. In S. Arieti (ed.), *American Handbook of Psychiatry*, Vol.3. (pp.183-200). New York: Basic Books.

- Rogers, C.R. (1980). *A Way of Being*. Boston, MA: Houghton Mifflin.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, 35(1), 113-130.
- Ryff, C. D. (1989). Happiness is everything, or is it – Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57, 1069–1081.
- Ryff, C.D. & Keyes, C.L.M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69, 719-727.
- Ryff, C. D., & Singer, B. (2008). Know thyself and become what you are: A eudaimonic approach to psychological well-being. *Journal of Happiness Studies*, 9, 13-39.
- Salovey, P., & Mayer, J. D.. (1990). Emotional intelligence. *Imagination, Cognition and Personality*, 9, 185–211.
- Sarıçam, H. (2015). Life satisfaction: Testing a structural equation model based on authenticity and subjective happiness. *Polish Psychological Bulletin*, 46(2), 278-284.
- Sartre, J-P. (1944). *Anti-Semite and Jew* (G.J. Becker, Trans.). New York: Schocken Books.
- Schmid, P. (2001). Authenticity: The person as his or her own author. In G. Wyatt (Ed.), *Rogers' Therapeutic Conditions: Evolution, Theory and Practice Volume 1: Congruence*. (pp. 213-228). Ross-on-Wye: PCCS Books.
- Schmid, P.F. (2005). Authenticity and alienation: Towards an understanding of the person beyond the categories of order and disorder. In S. Joseph & R. Worsley (Eds.), *Person-centred psychopathology: A positive psychology of mental health* (pp. 74-89). Ross-on-Wye: PCCS Books.

- Schneider, K. (2011). Toward a humanistic positive psychology: Why can't we just get along? *Self & Society*, 38, 18-25.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Sedikides, C., & Skowronski, J. J. (1997). The symbolic self in evolutionary context. *Personality and Social Psychology Review*, 1, 80-102.
- Sedikides, C., Slabu, L., Lenton, A.P., & Thomaes, S. (2017). State Authenticity. *Current Directions in Psychological Science*, 26(6), 521-525.
- Seligman, M.E.P. (1999). The president's address. *American Psychologist*, 54, 559-562.
- Seligman, M.E.P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, 55, 5-14.
- Seligman, M. E., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: empirical validation of interventions. *American psychologist*, 60(5), 410.
- Seto, E., Hicks, J. A., Vess, M., & Geraci, L. (2016). The association between vivid thoughts of death and authenticity. *Motivation and Emotion*, 40(4), 520-540.
- Shamsi, A., Ghamarani, A., Samadi, M., & Ahmadzadeh M. (2012). The Study Of The Validity And Reliability Of The Authentic Personality Scale. *Journal Of Psychological Models And Methods*, 2(8), 87-99.
- Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591-611.
- Sheldon, K. M. (2004). Integrity (honesty/authenticity). In C. Peterson & M. E. P. Seligman (Eds.), *Character strengths and virtues* (pp. 249- 272). New York: Oxford University Press.

- Sheldon, K. M., Ryan, R. M., Rawsthorne, L. A. I. R. D., & Ilardi, B. (1997). Trait self and true self: Cross-role variation in the Big Five traits and its relations with authenticity and subjective well-being. *Journal of Personality and Social Psychology, 73*(6), 1380-1393.
- Sherman, G. L. (2009). Martin Heidegger's Concept of Authenticity: A Philosophical Contribution to Student Affairs Theory. *Journal of College and Character, 10*(7), 1-8.
- Sijtsma, K. (2012). Future of psychometrics: Ask what psychometrics can do for psychology. *Psychometrika, 77*(1), 4-20.
- Slabu, L., Lenton, A. P., Sedikides, C., & Bruder, M. (2014). Trait and state authenticity across cultures. *Journal of Cross-Cultural Psychology, 45*(9), 1347-1373.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980) *Benefits of psychotherapy*. Baltimore, MD: Johns Hopkins University Press.
- Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice, 17*(2), 268–274.
- Spitzer, R.L., Kroenke, K., Williams, J.B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*, 1092-1097.
- Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology, 53*, 80–93.
- Steinmetz, G. (1998). Critical realism and historical sociology. A review article. *Comparative studies in society and history, 40*(01), 170-186.
- Stephen, S. & Elliott, R. (2017). The Strathclyde Inventory: Measuring congruence as an outcome for therapy. Paper presented at BACP Research Conference, Chester, UK.

- Sterne, J. A., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal*, 4(2), 127-141.
- Stevens, F. L. (2017). Authenticity: a mediator in the relationship between attachment style and affective functioning. *Counselling Psychology Quarterly*, 30(4), 392–414.
- Stratford, P. W. & Riddle, D. L. (2005). Assessing sensitivity to change: choosing the appropriate change coefficient. *Health and Quality of Life Outcomes*, 3, 23-30.
- Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, 80(3), 217-222.
- Tarescavage, A. M., & Ben-Porath, Y. S. (2014). Psychotherapeutic outcomes measures: a critical review for practitioners. *Journal of clinical psychology*, 70(9), 808-830.
- Tasca, G. A., & Gallop, R. (2009). Multilevel modeling of longitudinal data for psychotherapy researchers: I. The basics. *Psychotherapy Research*, 19(4-5), 429-437.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, 2, 53.
- Tennant, R., Hillier, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). The Warwick–Edinburgh Mental Well Being Scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes*, 5, 63.
- Thomaes, S., Sedikides, C., van den Bos, N., Hutteman, R., & Reijntjes, A. (2017). Happy To Be “Me?” Authenticity, Psychological Need Satisfaction, and Subjective Well-Being in Adolescence. *Child Development*, 88(4), 1045–1056.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.

- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58(1), 6-20.
- Vainio, M. M. & Daukantaite, D. (2016). Grit and Different Aspects of Well-Being: Direct and Indirect Relationships via Sense of Coherence and Authenticity. *Journal of Happiness Studies*, 17, 2119-2147.
- Van Dam, N. T., Hobkirk, A. L., Danoff-Burg, S., & Earleywine, M. (2012). Mind Your Words: Positive and Negative Items Create Method Effects on the Five Facet Mindfulness Questionnaire. *Assessment*, 19(2), 198–204.
- van den Bosch, R., & Taris, T. W. (2014). Authenticity at work: Development and validation of an individual authenticity measure at work. *Journal of Happiness Studies*, 15(1), 1-18.
- Vess, M., Leal, S. A., Hoeldtke, R. T., Schlegel, R. J., & Hicks, J. A. (2016). True self-alienation positively predicts reports of mindwandering. *Consciousness and Cognition*, 45, 89–99.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48. URL: <http://www.jstatsoft.org/v36/i03/>.
- Viera, A. J. & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 37(5), 360-363.
- Vogt, W. P. (2005). Discriminant validity. In W. P. Vogt (Ed.) *Dictionary of Statistic & Methodology* (3rd edn). Thousand Oaks: Sage Publications.
- von Glaserfeld, E. (1993). Learning and adaptation in the theory of constructivism. *Communication and Cognition*, 26, 393–402.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wampold, B. E. (2015). Routine outcome monitoring: Coming of age—With the usual developmental challenges. *Psychotherapy, 52*(4), 458-462.

Wampold, B. E., & Brown, G. S. J. (2005). Estimating variability in outcomes attributable to therapists: a naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology, 73*(5), 914-923.

Wampold, B. E. & Imel, Z. E. (2015) *The great psychotherapy debate. The evidence for what make psychotherapy work* (2nd edn). New York: Routledge/Taylor & Francis Group.

Wang, Y. N. (2015). Authenticity and relationship satisfaction: Two distinct ways of directing power to self-esteem. *PLoS One, 10*(12), e0146050.

Warwick Medical School (2018). WEMWBS in other languages. Retrieved March 9, 2019, from <https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/researchers/languages>

Waterman, A. S. (2013). The humanistic psychology–positive psychology divide: Contrasts in philosophical foundations. *American Psychologist, 68*(3), 124.

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 & Social Psychology, 6*(54), 1063-70.

Weir, K.F. & Jose, P. E. (2010). The perception of false self scale for adolescents: Reliability, validity, and longitudinal relationships with depressive and anxious symptoms. *British Journal of Developmental Psychology, 28*, 393-411.

- White, N. (2011). *An examination of dispositional authenticity* (Doctoral dissertation, Arizona State University).
- Winnicott, D. (1965). *The Maturation Process and the Facilitating Environment*, London: Hogarth Press.
- Wolpert, M., Jacob, J., Napoleone, E., Whale, A., Calderon, A., & Edbrooke-Childs, J. (2016). *Child- and Parent-reported Outcomes and Experience from Child and Young People's Mental Health Services 2011–2015*. London: CAMHS Press.
- Wood, A. M., Linley, P. A., Maltby, J., Baliousis, M., & Joseph, S. (2008). The authentic personality: A theoretical and empirical conceptualization and the development of the authenticity scale. *Journal of Counseling Psychology*, 55(3), 385–399.
- Wyatt, G. (2001). *Rogers' Therapeutic Conditions: Evolution, Theory and Practice Volume 1: Congruence*. Ross-on-Wye: PCCS Books.
- Yalom, I. D. (1980). *Existential psychotherapy*. New York: Basic Books.
- Young, R & Collin, A. (2004). Introduction: constructivism and social constructionism in the career field. *Journal of Vocational Behaviour* 64(3), 373-388.
- Zhang, C., Hirschi, A., Dik, B. J., Wei, J., & You, X. (2018). Reciprocal relation between authenticity and calling among Chinese university students: A latent change score approach. *Journal of Vocational Behavior*, 107, 222-232.
- Zimmerman, M., & McGlinchey, J. B. (2008). Why don't psychiatrists use scales to measure outcome when treating depressed patients? *Journal of Clinical Psychiatry*, 69, 1916-1919.
- Zung, W. W. (1971). A rating instrument for anxiety disorders. *Psychosomatics: Journal of Consultation and Liaison Psychiatry*, 12(6), 371-379.

Appendices

Appendix A: The Authenticity Scale

THE AUTHENTICITY SCALE

Directions: Below are a series of statements. Respond to each statement by writing the number from the scale below that you feel most accurately represents your response to the statement. There are no right or wrong responses, so please answer as honestly as you can.

1	2	3	4	5	6	7
Does not describe me at all						Describes me very well

SCORE

1. "I think it is better to be yourself, than to be popular."	
2. "I don't know how I really feel inside."	
3. "I am strongly influenced by the opinions of others."	
4. "I usually do what other people tell me to do."	
5. "I always feel I need to do what others expect me to do."	
6. "Other people influence me greatly."	
7. "I feel as if I don't know myself very well."	
8. "I always stand by what I believe in."	
9. "I am true to myself in most situations."	
10. "I feel out of touch with the 'real me.'"	
11. "I live in accordance with my values and beliefs."	
12. "I feel alienated from myself."	

To be completed by therapist:

SESSION NUMBER

THERAPIST CODE

PARTICIPANT CODE

Copyright © 2008 by The American Psychological Association. Reproduced with permission.

Appendix B: Permission from APA to use the Authenticity Scale

RightsLink Printable License

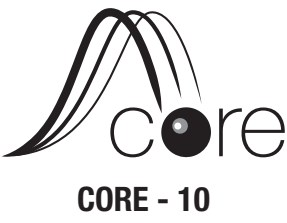
22/11/2017, 16:21

AMERICAN PSYCHOLOGICAL ASSOCIATION ORDER DETAILS

Nov 22, 2017

Order Number	501338653
Order date	Nov 22, 2017
Licensed Content Publisher	American Psychological Association
Licensed Content Publication	Journal of Counseling Psychology
Licensed Content Title	The authentic personality: A theoretical and empirical conceptualization and the development of the Authenticity Scale.
Licensed Content Author	Wood, Alex M.; Linley, P. Alex; Maltby, John; Baliousis, Michael; Joseph, Stephen
Licensed Content Date	Jul 1, 2008
Licensed Content Volume	55
Licensed Content Issue	3
Type of Use	Thesis/Dissertation
Requestor type	Academic institution
Format	Print, Electronic
Portion	Measure, scale or instrument
Rights for	Main product
Duration of use	
Creation of copies for the disabled	no
With minor editing privileges	yes
Distributing to	
In the following language(s)	Original language of publication
With incidental promotional use	no
The lifetime unit quantity of new product	0 to 499
The requesting person/organization is:	Daniel Morken / University of Roehampton
Order reference number	
Title of your thesis / dissertation	The Authenticity Scale as an outcome measure for psychological therapies - psychometrics and acceptability
Expected completion date	Nov 2018
Expected size (number of pages)	120
Requestor Location	University of Roehampton University of Roehampton Department of Psychology Holybourne Avenue London, SW154JD

Appendix C: CORE-10

	Site ID <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> Male <input type="checkbox"/>
	letters only <input type="text"/> <input type="text"/> numbers only <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Age <input type="text"/> <input type="text"/> Female <input type="checkbox"/>
Client ID Therapist ID <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> numbers only (1) <input type="text"/> <input type="text"/> numbers only (2) <input type="text"/> <input type="text"/>	Stage Completed S Screening R Referral A Assessment F First Therapy Session P Pre-therapy (unspecified) D During Therapy L Last Therapy Session X Follow up 1 Y Follow up 2	
Sub codes D D M M Y Y Y Y <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Stage <input type="text"/>	Episode <input type="text"/>
Date form given <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		

IMPORTANT – PLEASE READ THIS FIRST

This form has 10 statements about how you have been OVER THE LAST WEEK.
 Please read each statement and think how often you felt that way last week.
 Then tick the box which is closest to this.
Please use a dark pen (not pencil) and tick clearly within the boxes.

Over the last week		<i>Not at all</i> <i>Only Occasionally</i> <i>Sometimes</i> <i>Often</i> <i>Most or all the time</i>				
		0	1	2	3	4
1	I have felt tense, anxious or nervous	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	I have felt I have someone to turn to for support when needed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	I have felt able to cope when things go wrong	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Talking to people has felt too much for me	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	I have felt panic or terror	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	I made plans to end my life	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	I have had difficulty getting to sleep or staying asleep	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	I have felt despairing or hopeless	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	I have felt unhappy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	Unwanted images or memories have been distressing me	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Total (Clinical Score*)		<input style="width: 100px; height: 30px;" type="text"/>				

* **Procedure:** Add together the item scores, then divide by the number of questions completed to get the mean score, then multiply by 10 to get the Clinical Score.
Quick method for the CORE-10 (if all items completed): Add together the item scores to get the Clinical Score.

THANK YOU FOR YOUR TIME IN COMPLETING THIS QUESTIONNAIRE

Appendix D: Ethics application approval

Ethics Application

Applicant: Daniel Morken

Title: The Authenticity Scale as an outcome measure for psychological therapies - psychometrics and acceptability

Reference: PSYC 17/ 275

Department: Psychology

Many thanks for your response and the amended documents. Under the procedures agreed by the University Ethics Committee I am pleased to advise you that your Department has confirmed that all conditions for approval of this project have now been met (but please see the minor conditions below). We do not require anything further in relation to this application.

Minor Conditions:

- i. Thank you for deleting your personal mobile number from relevant documentation. If there is an alternative office or work-related phone number that could be used instead (e.g. DoS number), please do insert.
- ii. Please let us have a copy of the final version of the Therapy Helpful Survey if it changes from the one submitted.

As these are only minor conditions it is assumed that you will adhere to these conditions for approval and therefore we do not require a response.

Please note that on a standalone page or appendix the following phrase should be included in your thesis:

The research for this project was submitted for ethics consideration under the reference PSYC 17/ 275 in the Department of Psychology and was approved under the procedures of the University of Roehampton's Ethics Committee on 18.09.17.

Please Note:

- **This email confirms that all conditions have been met and thus confirms final ethics approval (it is assumed that you will adhere to any minor conditions still outstanding, therefore we do not require a response to these).**
- **University of Roehampton ethics approval will always be subject to compliance with the University policies and procedures applying at the time when the work takes place. It is your responsibility to ensure that you are familiar and compliant with all such policies and procedures when undertaking your research.**
- **Please advise us if there are any changes to the research during the life of the project. Minor changes can be advised using the Minor Amendments Form on the Ethics Website, but substantial changes may require a new application to be submitted.**
- **If this project involves clinical procedures or administering substances it is a condition of Ethics approval that all relevant SOPs published on the department communities pages are fully complied with.**

Many thanks,

Jan

Jan Harrison
Ethics Officer
Research Office
University of Roehampton | London | SW15 5PJ
jan.harrison@roehampton.ac.uk | www.roehampton.ac.uk
Tel: +44 (0) 20 8392 5785

Consider the environment. Please don't print this e-mail unless you really need to.

Appendix E: Recruitment email

Dear ...,

My name is Daniel Morken, I'm a trainee counselling psychologist at the University of Roehampton.

My doctoral thesis, supervised by Prof. Mick Cooper, is on validating a person-centered measure of authenticity, based on Rogers' writings on congruence, as an outcome measure for counselling and psychotherapy. We're writing this email to ask if you might be interested in participating in our research.

Participating would involve asking your clients who are 18 years and above to complete 'the Authenticity Scale' along with the CORE-10, at the beginning of therapy, at the end of their therapy and on every 6th session. If you are already using an outcome measure, we could likely use that as the second measure, reducing how many forms participating clients would need to complete. On the final session, participants would also be asked to complete the Therapy Questionnaire Helpfulness Survey, a short questionnaire asking clients about how they found the questionnaires they completed.

Our study is in collaboration with Prof. Stephen Joseph, one of the original developers of the Authenticity Scale. The scale is a 12-item measure, based closely on the person-centred definition of congruence. Our rationale behind this study is that person-centred theory holds increased authenticity to be the ultimate aim of psychotherapy. Thus, it would make sense to also measure the effectiveness of therapy with a measure of authenticity. In addition, the current emphasis on evidence-based practice in much of the therapy world, brings an increasing pressure to use outcome measures to document effectiveness. However, as the most widely used outcome measures tend to be symptom and distress focused, measuring the extent to which the client is suffering. With this research we're aiming to help answer the call for outcome measures that instead focus on wellbeing, measuring the effectiveness of therapy through the increase in authenticity.

If you would be interested in participating in our study, or would like more info, please do not hesitate to contact me on:

morkend@roehampton.ac.uk

All the best,

Daniel

Appendix F: Counselling service information sheet



COUNSELLING SERVICE INFORMATION SHEET *The Authenticity Scale as an outcome measure for therapy*

Thank you for reading this information sheet. This document will explain why we are doing this research, and set out what will be involved for counselling services who choose to take part. We appreciate you taking the time to read it, and hope you will be interested in participating in this research.

The Research Project

This research aims to explore whether the 'Authenticity Scale', an outcome measure based on person-centred theory and positive psychology, would be a useful measure for evaluating the effectiveness of counselling and psychotherapy. Considering the current emphasis on evidence-based practice in much of the therapy world, there is increasing pressure to use outcome measures to document effectiveness. However, the most widely used outcome measures tend to be symptom and distress focused, that is, they measure the extent to which the client is suffering. This research is aiming to help answer the call for outcome measures that instead focus on wellbeing, measuring the effectiveness of therapy through the increase in authenticity, as is key to several psychological therapies, including humanistic and psychodynamic.

Research procedure

This research is looking to have adults (18+) who are receiving therapy complete two short questionnaires, the Authenticity Scale and the CORE-10 (though if your service already use a different outcome measures to the CORE-10, we can use that as the second measure). These two questionnaires have 22 questions altogether, and would need to be completed on the first session, on the final session and on every 6th session. On the 12th session (or the final session, if you are doing less than 12) there would be an additional short questionnaire (10 questions) asking clients how helpful they found the two questionnaires they have completed. In addition, clients will also be asked to complete a basic demographics sheet.

Questionnaire packs will be given to the client's counsellor who will give them to the client to complete. Completed questionnaires can be discussed in-session.

Counselling services' involvement

Counselling services interested in supporting this research will be asked to give their therapists interested in participating the Therapist Information Sheet and Consent Form. Therapists can then give their adult clients about to start counselling the Client Information Sheet and the Client Consent Form, either at point of contact with the service, or during an initial assessment session. Clients can take the Client Information Sheet away with them to read, and can sign the consent form before their first session if they wish to participate.

Consent

Consent to participate in this research will first be obtained from the counselling service, then from the individual therapists and finally from participating clients.

Potential disadvantages/ risks to participants

A potential disadvantage of participating in this research is that clients will have to spend 5 minutes completing questionnaires and might find that boring. There are no expected risks for clients who take part in the study. For the therapist, participating would involve giving their clients the questionnaire pack on the first and final sessions, as well as on every 6th session, but otherwise there would be no disadvantages.

Potential benefits to participating

There is no direct benefit to taking part in this study, although the questionnaires can provide useful information concerning how the client is doing. The information gathered from this project will contribute to the research regarding the use of outcome measures that are not distress or symptom focused, hopefully providing therapists with an alternative approach to evaluating the effectiveness of therapy.

Confidentiality, anonymity and data storage

All information provided will be kept confidential, and only accessible to members of the research team. All collection, storage and processing of data will comply with the principles of the Data Protection Act 1998, and has been approved under the procedures of the University of Roehampton Ethics Committee. All the information provided will be stored securely and, where possible, anonymized. All data included in the publication or presentation of this research, and any subsequent research publications, will be fully anonymised to ensure that no individual is identifiable.

Completed questionnaires will be stored physically separate from signed consent forms, and there will be no way to match consent forms to completed questionnaires. Anonymised data will be stored for an indefinite period of time following the study (for at least 10 years), and may be used for publication, presentation, or for subsequent research projects or data analyses. Consent forms will be destroyed after ten years.

Withdrawal

If clients wish to withdraw from the study, they can at any point message the Chief Investigator with their unique ID code. However, data may still be used in aggregate form.

Dissemination of findings

The results of this research study will be written up in fulfilment of the requirements for the Doctorate in Counselling Psychology from the University of Roehampton. The results of this research may be published in academic journals, or presented at conferences.

Who is organising the research?

This research is being undertaken by the Department of Psychology at the University of Roehampton. This project has been approved under the procedures of the University of Roehampton's Ethics Committee.

What happens now?

If you have decided you would like to participate in this research, and you have signed the Counselling Service Consent Form, the next thing to do is to give your therapists the Therapist Information Sheet and Therapist Consent Form. Participating therapists will then give any adult clients (18+) the Client Information Sheet and Consent Form prior to their starting therapy.

.....
If you have any further questions about any aspect of your participation, please contact Daniel Morken (chief investigator) for more details. However, if you would like to contact an independent party please contact the Head of Department:

Chief Investigator:
Daniel Morken
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
morkend@roehampton.ac.uk

Director of Studies:
Prof. Mick Cooper
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
mick.cooper@roehampton.ac.uk
0208 392 3741

Head of Department:
Dr. Diane Bray
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
d.bray@roehampton.ac.uk
0208 392 3741

Appendix G: Counselling service consent form



COUNSELLING SERVICE CONSENT FORM

The Authenticity Scale as an outcome measure for therapy

Consent Statement:

I have read the Counselling Service Information Sheet and understand the purpose and procedure of this research. I understand that I may request further details and information should I wish. On behalf of the counselling service I agree to take part in this research, and I am aware that our participation is entirely voluntary. I understand that I am free to withdraw any client receiving counselling at our service at any stage in the proceedings and also to withdraw from the project altogether without giving a reason, though collected data may still be used in aggregate form. I consent to participating clients completing outcome measures, and for the data collected to be used in the preparation of a thesis and accompanying papers and presentations. I understand that anonymised data will be stored indefinitely (for at least 10 years) and that this data may be used for other research projects and data analyses (at the discretion of the investigator). I understand that the information participants provide will be treated confidentially by the investigator and research team. I understand that data cannot be traced back to participants' identities, and the identity of the counselling service will be removed in the publication or presentation of any findings. I understand that data will be collected and processed in accordance with the Data Protection Act 1998 and with the University of Roehampton's Data Protection Policy.

Name

Organisation

Signature

Date

If you have any further questions, please contact Daniel Morken (chief investigator) for more details:

Chief Investigator

Daniel Morken
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
morkend@roehampton.ac.uk

Please note: if you have a concern about any aspect of your participation or any other queries please raise this with the investigator. However, if you would like to contact an independent party please contact the Head of Department.

Director of Studies Contact Details:

Prof. Mick Cooper
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
mick.cooper@roehampton.ac.uk
0208 392 3741

Head of Department Contact Details:

Dr. Diane Bray
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
d.bray@roehampton.ac.uk
0208 392 3741

Appendix H: Therapist information sheet



University of
Roehampton
London

THERAPIST INFORMATION SHEET

The Authenticity Scale as an outcome measure for therapy

Thank you for reading this information sheet. This document will explain why we are doing this research and set out what will be involved for therapists who choose to take part. We appreciate you taking the time to read it and hope you will be interested in participating in this research.

The Research Project

This research aims to explore whether the Authenticity Scale, an outcome measure based on person-centred theory and positive psychology, would be a useful measure for evaluating the effectiveness of counselling and psychotherapy. Considering the current emphasis on evidence-based practice in much of the therapy world, there is increasing pressure to use outcome measures to document effectiveness, which in turn can help with funding. However, the most widely used outcome measures tend to be symptom and distress focused, that is, they measure the extent to which the client is suffering. This research is aiming to help answer the call for outcome measures that instead focus on wellbeing, measuring the effectiveness of therapy through the increase in authenticity, as is key to several psychological therapies, including humanistic and psychodynamic.

Research procedure

This research is looking to have adults (18+) who are receiving therapy complete two short questionnaires, the Authenticity Scale and the CORE-10 (though if your service already use a different outcome measures to the CORE-10, we can use that as the second measure). These two questionnaires have 22 questions altogether, and would need to be completed on the first session, on the final session and on every 6th session. On the 12th session (or the final session, if you are doing less than 12) there would be an additional short questionnaire (10 questions) asking clients how helpful they found the two questionnaires they have completed. In addition, clients will also be asked to complete a basic demographics sheet.

Therapists' involvement

For therapists interested in supporting this research, you will be provided with Client Consent forms and information sheets by the Chief Investigator, as well as questionnaire packs. You will be asked to give clients meeting the inclusion criteria the Client Information Sheet and Consent Form prior to their starting therapy, either at point of contact with the service, or during the initial assessment session. Clients can then take the Client Information Sheet away with them to read, and can sign the consent form before their first session if they wish to participate.

Therapists will be asked to give participating clients the questionnaire packs on the first session, the and on the final session, as well as on every 6th session (6th, 12th, 18th etc). Questionnaires will preferably be completed before the start of the session.

Therapists willing to participate will also be asked to complete a Therapist Questionnaire Form, asking about basic demographics, how long you have been practicing, your level of training and your therapeutic approach. This form will be anonymous and confidential.

It is important to be aware that question 6 on the CORE-10 ('I have made plans to end my life.') can be an important indicator of risk for the client, and as such relevant for safeguarding. The Information Sheet given to clients makes it clear that therapists may see the completed questionnaires.

Consent

Consent to participate in this research will first be obtained from the counselling service, then from the individual therapists and finally from participating clients.

Potential disadvantages/ risks to participants

A potential disadvantage of participating in this research is that clients will have to spend 5 minutes to complete questionnaires and might find that boring. There are no expected risks for clients who take part in the study. For the therapist, participating would involve giving your clients the questionnaire pack on the first, and final sessions, as well as on every 6th session, but otherwise there would be no disadvantages.

Potential benefits to participating therapists

There is no direct benefit to taking part in this study, although the questionnaires can provide useful information concerning how the client is doing. The information gathered from this project will contribute to the research regarding the use of outcome measures that are not distress or symptom focused, hopefully providing therapists with an alternative approach to evaluating the effectiveness of therapy.

Confidentiality, anonymity and data storage

All data generated from this study will be stored securely to the highest possible standard of confidentiality. All collection, storage and processing of data will comply with the principles of the Data Protection Act 1998, and has been approved under the procedures of the University of Roehampton Ethics Committee. Completed questionnaires will be stored physically separate from signed consent forms, and there will be no way to match consent forms to completed questionnaires. If participants wish to withdraw from the study, they can at any point message the Chief Investigator with their unique ID code. However, data may still be used in aggregate form. Anonymised data generated from this study will be stored for an indefinite period of time following the study (for at least 10 years), and may be used for publication, presentation, or for subsequent research projects or data analyses. Consent forms will be destroyed after ten years. All data included in the publication or presentation of this research, and any subsequent research publications, will be fully anonymised to ensure that no individual is identifiable.

Dissemination of findings

The results of this research study will be written up in fulfilment of the requirements for the Doctorate in Counselling Psychology from the University of Roehampton. The results of this research may be published in academic journals, or presented at conferences.

Who is organising the research?

This research is being undertaken by the Department of Psychology at the University of Roehampton. This project has been approved under the procedures of the University of Roehampton's Ethics Committee.

What happens now?

If you have decided you would like to participate in this research, the next thing to do is to sign the Therapist Consent Form and give it to your service manager. Then, when you are to start working with a client who is 18 or older, please give them the Client Information Sheet and the Client Consent Form so they can decide if they wish to participate. Ideally give them these forms on an assessment session so that they can decide whether they wish to participate before their first session. Clients can then return the signed consent form on their first session and complete the first questionnaires.

.....
If you have any further questions about any aspect of your participation, please contact Daniel Morken (chief investigator) for more details. However, if you would like to contact an independent party please contact the Head of Department.:

Chief Investigator:
Daniel Morken
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
morkend@roehampton.ac.uk

Director of Studies:
Prof. Mick Cooper
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
mick.cooper@roehampton.ac.uk
0208 392 3741

Head of Department:
Dr. Diane Bray
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
d.bray@roehampton.ac.uk
0208 392 3741

Appendix I: Therapist consent form



THERAPIST CONSENT FORM

The Authenticity Scale as an outcome measure for therapy

Consent Statement:

I have read the Therapist Information Sheet and understand the purpose and procedure of this research. I understand that I may request further details and information should I wish. I agree to take part in this research, and I am aware that my participation is entirely voluntary. I understand that I am free to withdraw from the project at any stage in the proceedings, without giving a reason. I consent to completing the Therapist Questionnaire Form, and for the data collected to be used in the preparation of a doctoral thesis and accompanying papers and presentations. I understand that anonymised data will be stored indefinitely (for at least 10 years) and that this data may be used for other research projects and data analyses (at the discretion of the investigator). I understand that the information participants and therapists provide will be treated confidentially by the investigator and research team. I understand that data cannot be traced back to participants' identities, and the identity of the counselling service will be removed in the publication or presentation of any findings. I understand that data will be collected and processed in accordance with the Data Protection Act 1998 and with the University of Roehampton's Data Protection Policy.

Name

Organisation

Signature

Date

If you have any further questions, please contact Daniel Morken (chief investigator) for more details:

Chief Investigator

Daniel Morken
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
morkend@roehampton.ac.uk

Please note: if you have a concern about any aspect of your participation or any other queries please raise this with the investigator. However, if you would like to contact an independent party please contact the Head of Department.

Director of Studies Contact Details:

Prof. Mick Cooper
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
mick.cooper@roehampton.ac.uk
0208 392 3741

Head of Department Contact Details:

Dr. Diane Bray
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
d.bray@roehampton.ac.uk
0208 392 3741

Appendix J: Therapist demographics form



Therapist Code

Therapist Questionnaire Form

Please answer all questions you feel comfortable answering.

Therapist Demographics

Date of completion
DD MM YYYY

What is your age? _____

What is your gender? _____

What is your sexuality? _____

What is your ethnic group? White

- English/Welsh/Scottish/Northern Irish/British
- Gypsy or Irish Traveller
- Irish
- Any other white background, describe _____

Mixed/Multiple ethnic groups

- White and Black Caribbean
- White and Asian
- White and Black Asian
- Any other Mixed/Multiple ethnic background, describe _____

Asian/Asian British

- Indian
- Bangladeshi
- Pakistani
- Chinese
- Any other Asian background, describe _____

Black/African/Caribbean/Black British

- African
- Any other Black/African/Caribbean background, describe _____
- Caribbean

Other ethnic group

- Arab
- Any other ethnic group, describe _____

Do you have a disability?

- No
- Yes

If yes, please specify _____

Your Training and Practice

When did you qualify as a professional therapist?

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
MM		YYYY			

If you are not yet qualified, when do you expect to qualify?

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
MM		YYYY			

At what level was your counsellor/psychotherapy training programme?

- | | |
|--|--|
| <input type="checkbox"/> Level 5 (Diploma, Foundation, HE/FE etc.) | <input type="checkbox"/> Level 7 (Master's degree, PGDip etc.) |
| <input type="checkbox"/> Level 6 (Bachelor's degree) | <input type="checkbox"/> Level 8 (Doctoral degree) |
| <input type="checkbox"/> Other describe _____ | |

What was the orientation of the training? (please tick one or more)

- | | | |
|---|--|--------------------------------------|
| <input type="checkbox"/> Humanistic | <input type="checkbox"/> Behavioural | <input type="checkbox"/> Existential |
| <input type="checkbox"/> Person-centred | <input type="checkbox"/> Systemic | <input type="checkbox"/> Integrative |
| <input type="checkbox"/> Gestalt | <input type="checkbox"/> Cognitive-behavioural | <input type="checkbox"/> Eclectic |
| <input type="checkbox"/> Transactional analysis | <input type="checkbox"/> Cognitive | <input type="checkbox"/> Pluralistic |
| <input type="checkbox"/> Psychodynamic/analytic | | |
| <input type="checkbox"/> Other specify _____ | | |

What is your principal professional identity?

- | | |
|--|---|
| <input type="checkbox"/> Counsellor | <input type="checkbox"/> Counselling Psychologist |
| <input type="checkbox"/> Psychotherapist | <input type="checkbox"/> Clinical Psychologist |
| <input type="checkbox"/> Other specify _____ | |

What is your principal therapeutic orientation? (please tick one or more)

- | | | |
|---|--|--------------------------------------|
| <input type="checkbox"/> Humanistic | <input type="checkbox"/> Behavioural | <input type="checkbox"/> Existential |
| <input type="checkbox"/> Person-centred | <input type="checkbox"/> Systemic | <input type="checkbox"/> Integrative |
| <input type="checkbox"/> Gestalt | <input type="checkbox"/> Cognitive-behavioural | <input type="checkbox"/> Eclectic |
| <input type="checkbox"/> Transactional analysis | <input type="checkbox"/> Cognitive | <input type="checkbox"/> Pluralistic |
| <input type="checkbox"/> Psychodynamic/analytic | | |
| <input type="checkbox"/> Other Specify _____ | | |

Appendix K: Client information sheet



CLIENT INFORMATION SHEET (please retain this form for your personal records)

The Authenticity Scale as an outcome measure for therapy

Thanks for reading this information sheet about our research.

We would like to invite you to take part in our research project. We are exploring whether the 'Authenticity Scale' is a useful questionnaire with which to evaluate how effective therapy is. Having good ways of evaluating how useful therapy is to people is important in making therapy even more beneficial and accessible, and so we would love your help.

Before you decide if you want to join in, it's important to understand why we are doing this research and what taking part involves. If you have any questions, feel free to contact the Chief Investigator whose details are at the bottom of this form.

Why are we doing this?

To provide the best possible therapy service to people who need it, it's important to understand how therapy works and what sorts of therapy are most helpful. One of the ways we can do this is by asking people who have therapy to complete a questionnaire that tries to measure how they are feeling before and after therapy and see if there has been any change. There are many questionnaires like this, but most of them focus on how much people are struggling. This research is investigating whether a questionnaire that focuses on wellbeing, might be a good way of seeing how useful therapy is. The Authenticity Scale is such a wellbeing focused questionnaire. It has 12 questions, all related to how 'in-touch' with yourself you feel.

Do I have to take part?

No, participation in this research is voluntary. If you consent to taking part, but later change your mind, you can still leave the study at any point without providing a reason (though questionnaires you've completed may still be used in compiled form) by contacting your counselling service. Taking part or choosing not to do so will in no way affect the therapy service you receive.

Why have I been asked to take part?

You have been asked to take part in this study because you are over 18, and you are about to start therapy at a therapy service that is participating in this research.

What would taking part mean?

Taking part in this study would mean completing two short questionnaires on your first therapy session, and on your final session, as well as on every 6th session. So, if you have 12 sessions, it would mean completing two questionnaires three times (Session 1, session 6 and session 12). These questionnaires have 10 and 12 questions, and should together take less than 5 minutes to do. There will be one additional short questionnaire to complete on your 12th session (or final session, if you are doing less than 12) (10 questions) asking you how helpful you found the other two questionnaires. There would also be a short, anonymous demographics sheet to complete on your first session.

Consent

To take part you would have to sign a Consent Form saying you have read and understood this information sheet and that you agree to participate in this project.

What are the advantages and disadvantages to taking part?

You would have to give up about 5 minutes to complete the questionnaires before your therapy session, and the questions might be boring, but other than that there are no disadvantages to taking part. On the other hand, a lot of people find that completing questionnaires like this can be really interesting and helpful for their therapy experience.

Can people trace my questionnaires back to me?

No, the questionnaires are completely anonymous and are only marked by a unique participant code to ensure this. However, your therapist and their supervisor might see your questionnaires and use this to make the therapy better for you. If you indicate on one of the questionnaires that you might be in danger or at risk, your therapist might have to break confidentiality according to the safeguarding procedures of the service where you are getting therapy.

The completed anonymous questionnaires will be stored securely and confidentially, only accessible to the research team. All collection, storage and processing of data will comply with the principles of the Data Protection Act 1998, and has been approved under the procedures of the University of Roehampton's Ethics Committee. Signed consent forms will be stored for 10 years, and the destroyed. Anonymised data generated from this project will be stored indefinitely (for at least 10 years).

What will happen to the results of the questionnaires?

The questionnaires will all be compiled, analysed and written up in a research report that might be published, in fulfilment for the doctorate in Counselling Psychology from the University of Roehampton. The results of this research may be published in academic journals, presented at conferences or used for teaching purposes. Results may also be used for other research studies, publications, presentations or subsequent projects (if the investigator gives permission for this).

Who is organising the research?

This research is being undertaken by the Department of Psychology at the University of Roehampton. This project has been approved under the procedures of the University of Roehampton's Ethics Committee.

What happens next?

If you have read this information sheet and decided you would like to participate in this research, then the next thing to do is to read and sign the Client Consent Form and bring it to your next therapy session. At this session your therapist will give you the two questionnaires to complete as well as an anonymous demographics sheet (asking about your age, gender and background).

If you do not want to take part there is nothing further you need to do.

.....
If you have any further questions, please contact Daniel Morken (chief investigator) for more details:

Chief Investigator

Daniel Morken
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
morkend@roehampton.ac.uk

Please note: if you have a concern about any aspect of your participation or any other queries please raise this with the investigator. However, if you would like to contact an independent party please contact the Head of Department.

Director of Studies Contact Details:

Prof. Mick Cooper
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
mick.cooper@roehampton.ac.uk
0208 392 3741

Head of Department Contact Details:

Dr. Diane Bray
Holybourne Avenue
Department of Psychology
University of Roehampton
London SW15 4JD
d.bray@roehampton.ac.uk
0208 392 3741

Appendix L: Client consent form



CLIENT CONSENT FORM

The Authenticity Scale as an outcome measure for therapy

Consent statement (please tick if you agree)

1.	I have read and understood the Information Sheet about the study	<input type="checkbox"/>
2.	I know I can contact the Chief Investigator with any questions about the study if I need to.	<input type="checkbox"/>
3.	I am 18 years old or older	<input type="checkbox"/>
4.	I agree to take part in the project. I understand that I am a volunteer.	<input type="checkbox"/>
5.	I understand I can leave the study at any time without giving a reason. I understand that leaving the study will not affect any therapy I receive. I understand that data may still be used in collated form.	<input type="checkbox"/>
6.	I have read and understood the part of the Information sheet about the use of the data in research, publications, presentations, sharing and storage.	<input type="checkbox"/>
7.	I understand that data will be collected and processed in accordance with the Data Protection Act 1998 and with the University of Roehampton's Data Protection Policy.	<input type="checkbox"/>
8.	I understand that my identity will be protected in any write-ups or articles of this study	<input type="checkbox"/>
9.	I understand that anonymised completed questionnaires will be store indefinitely (for at least 10 years). I understand that other researchers may be able to use the data if they agree to treat it confidentially (and with the permission of the researcher).	<input type="checkbox"/>
10.	I agree to sign and date this consent form.	<input type="checkbox"/>

Name

Signature

Date

Please complete and return this form to your therapist.

Chief Investigator Contact Details:

Daniel Morken
 Holybourne Avenue
 Department of Psychology
 University of Roehampton
 London SW15 4JD
morkend@roehampton.ac.uk

Please note: if you are worried about any aspect of your participation or any other queries please raise this with the investigator (or if the researcher is a student you can also contact the Director of Studies.) However, if you would like to contact an independent party please contact the Head of Department.

Director of Studies Contact Details:

Prof. Mick Cooper
 Holybourne Avenue
 Department of Psychology
 University of Roehampton
 London SW15 4JD
mick.cooper@roehampton.ac.uk
 0208 392 3741

Head of Department Contact Details:

Dr. Diane Bray
 Holybourne Avenue
 Department of Psychology
 University of Roehampton
 London SW15 4JD
d.bray@roehampton.ac.uk
 0208 392 3741

Appendix M: Client demographics form



Participant Code

Therapist Code

Demographics Form

Thank you for agreeing to take part in this research. Please answer all questions you feel comfortable answering.

- 1) What is your gender? _____
- 2) What is your sexuality? _____
- 3) What is your age? _____
-

4) What is your ethnic group?

White

- English/Welsh/Scottish/Northern Irish/British
- Irish
- Gypsy or Irish Traveller
- Any other white background, describe: _____

Mixed/Multiple ethnic groups

- White and Black Caribbean
- White and Black Asian
- White and Asian
- Any other Mixed/Multiple ethnic background, describe: _____

Asian/Asian British

- Indian
- Pakistani
- Bangladeshi
- Chinese
- Any other Asian background, describe: _____

Black/African/Caribbean/Black British

- African
- Caribbean
- Any other Black/African/Caribbean background, describe: _____

Other ethnic group

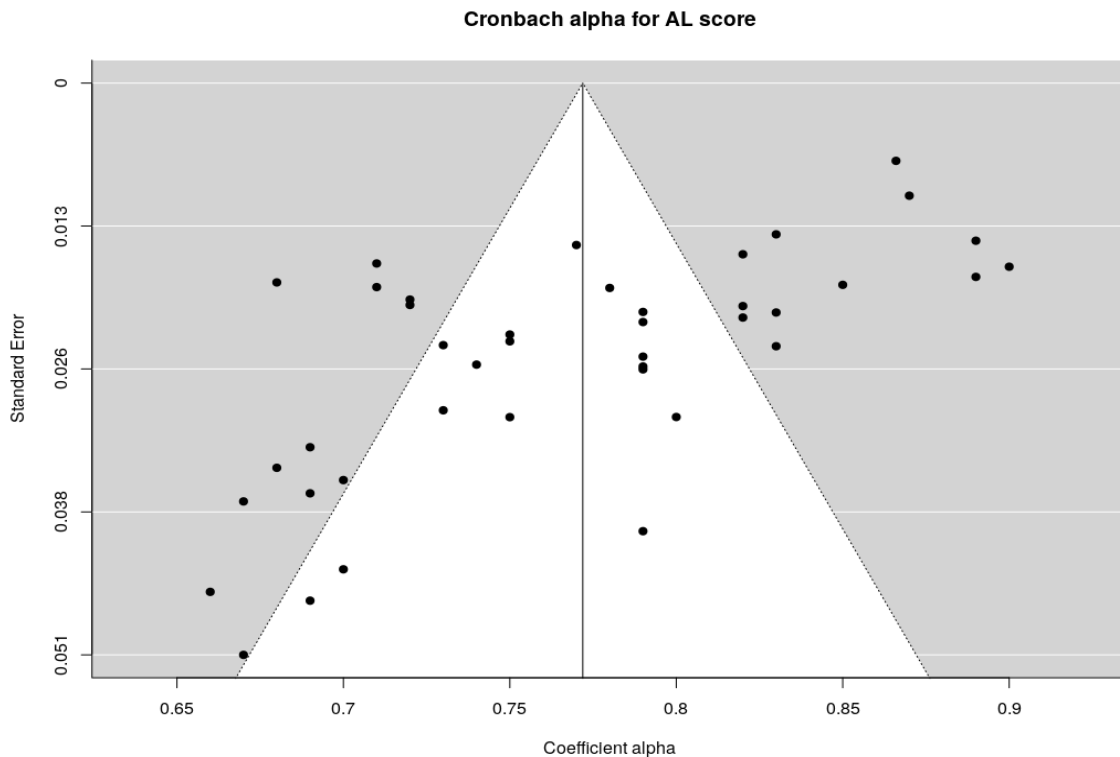
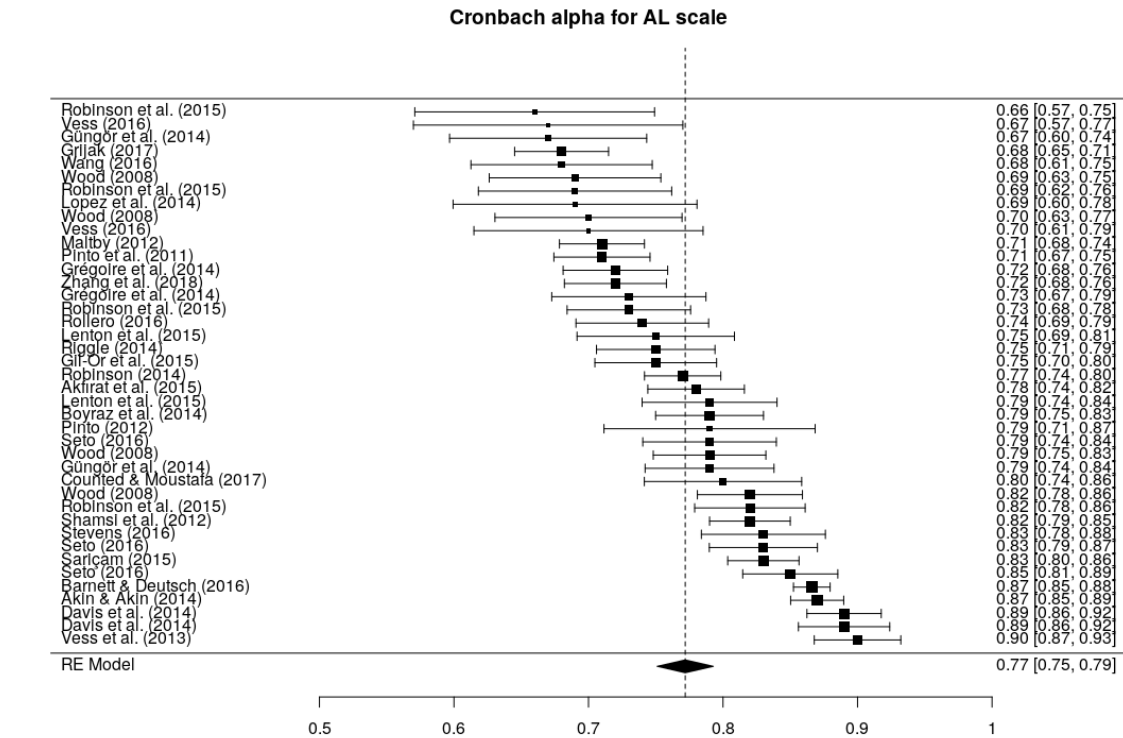
- Arab
- Any other ethnic group, describe: _____

5) What is your highest educational qualification?

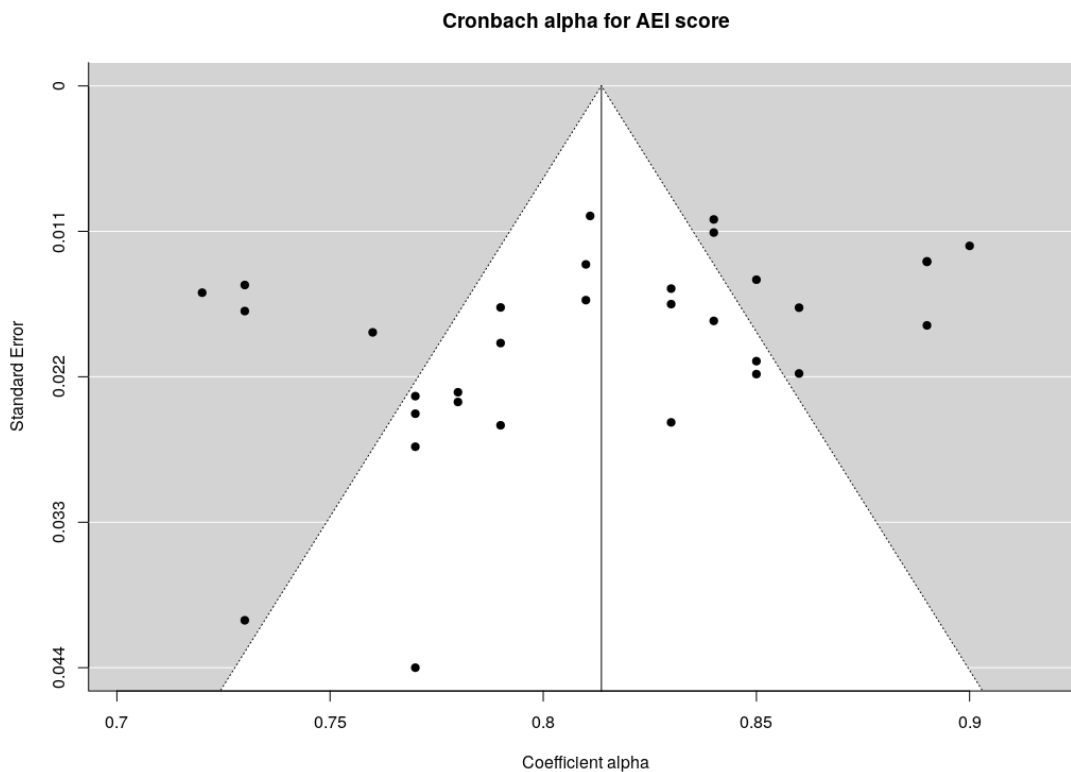
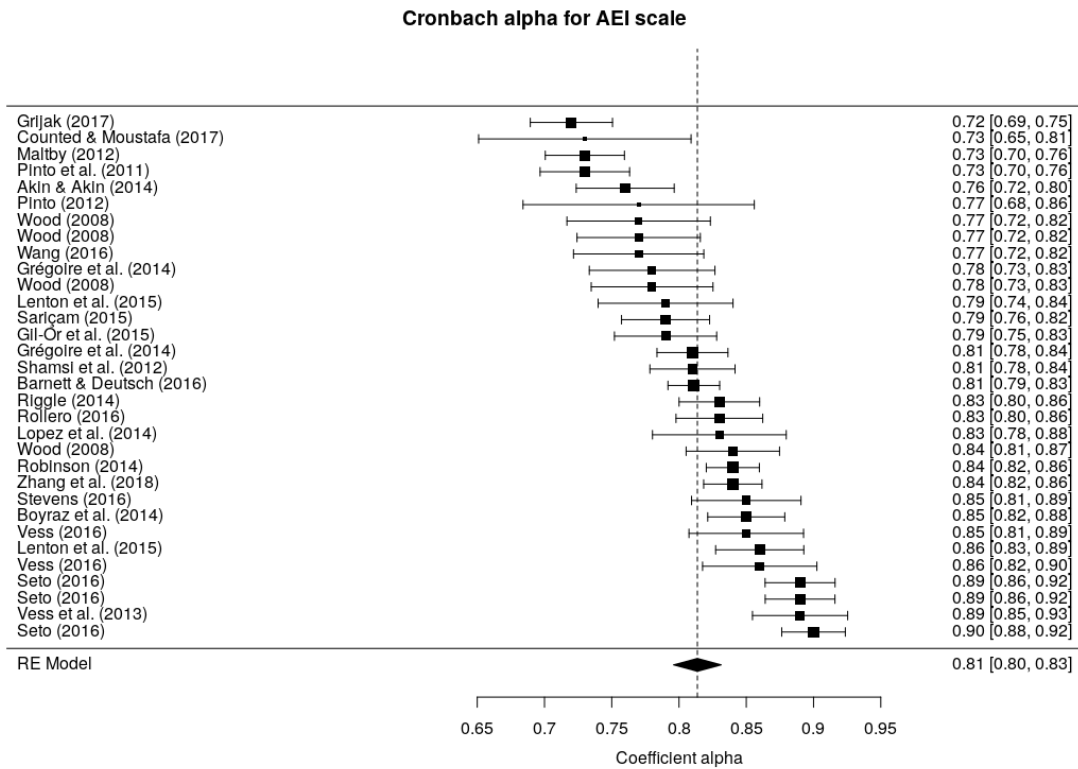
- Secondary School (GCSE/O-Levels)
- Post-Secondary School (College, A-Levels, NVQ3 or below, etc.)
- Vocational Qualification (Diploma, Certificate, Foundation, BTEC, NVQ4 and above, or similar)
- Undergraduate Degree (BA, BSc or similar.)
- Post-graduate Degree (MA, MSc, PGDip or similar.)
- Doctorate (PhD, MD, PsychD or similar.)
- Any other level, please describe: _____

Appendix N: Forest and funnel plots of the coefficient alphas for the AS subscales

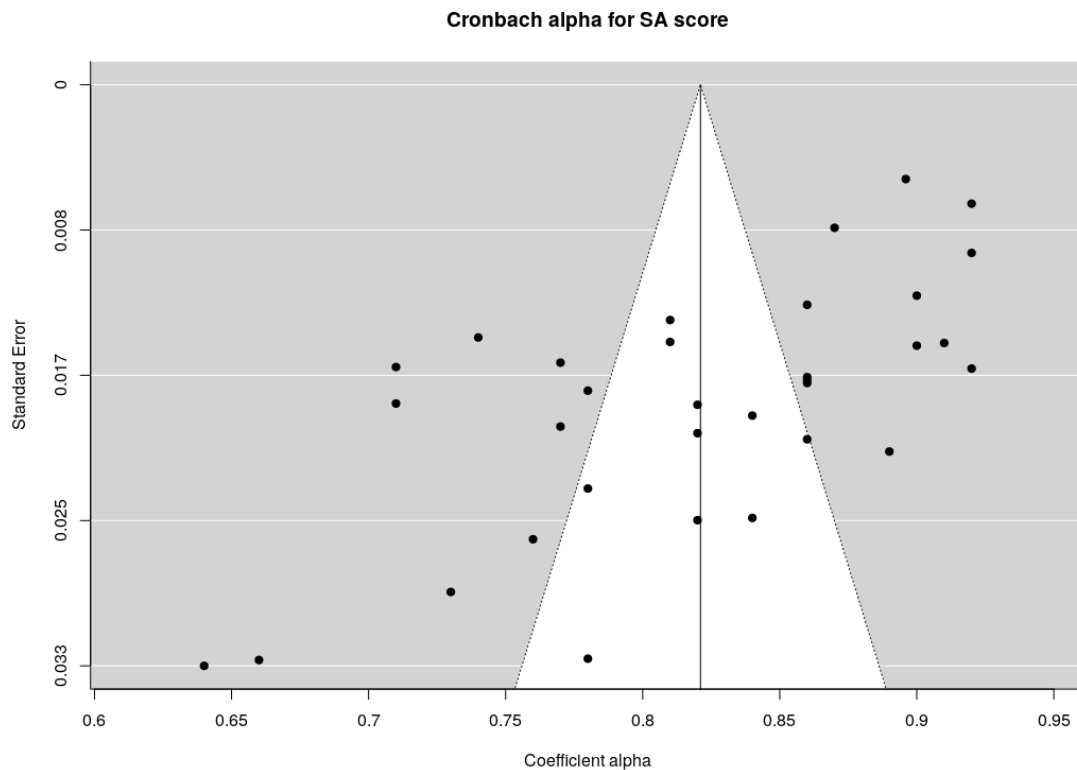
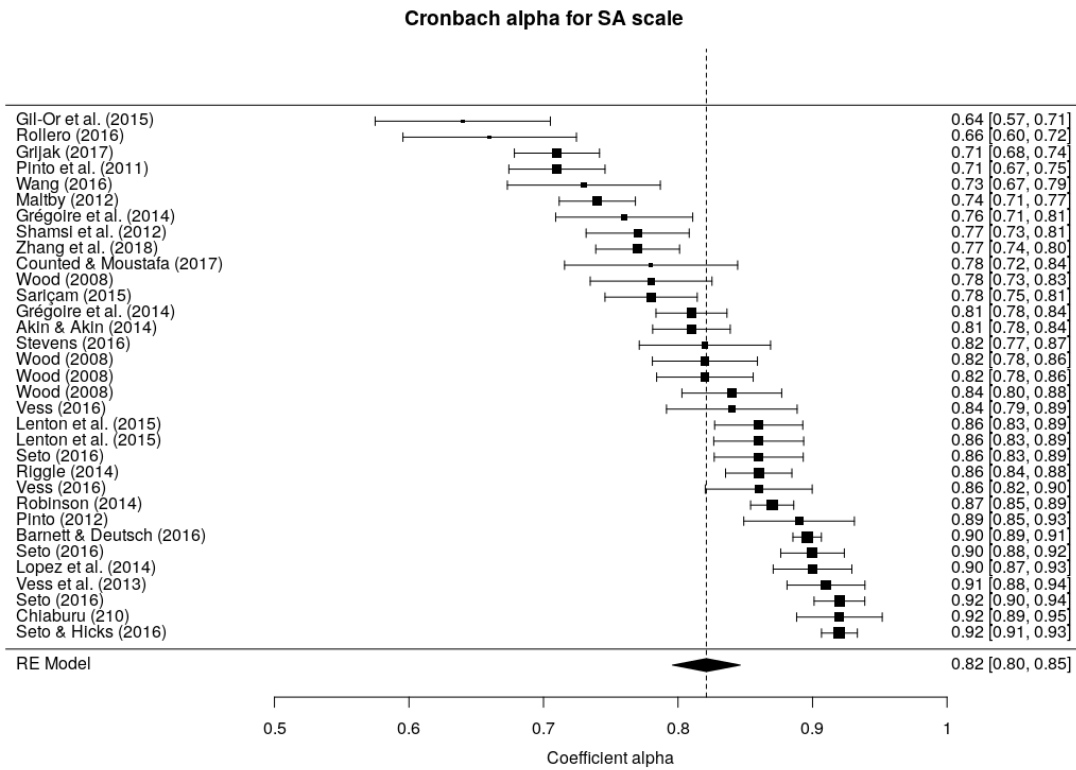
Authentic living subscale:



Accepting external influences subscale:

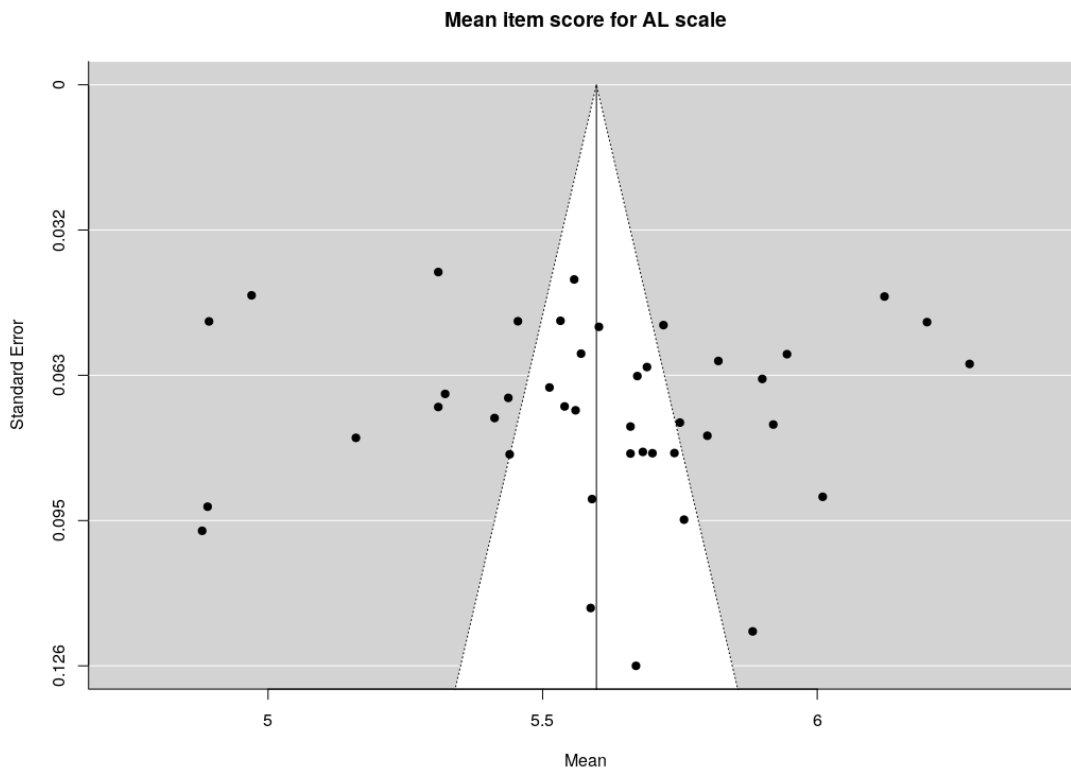
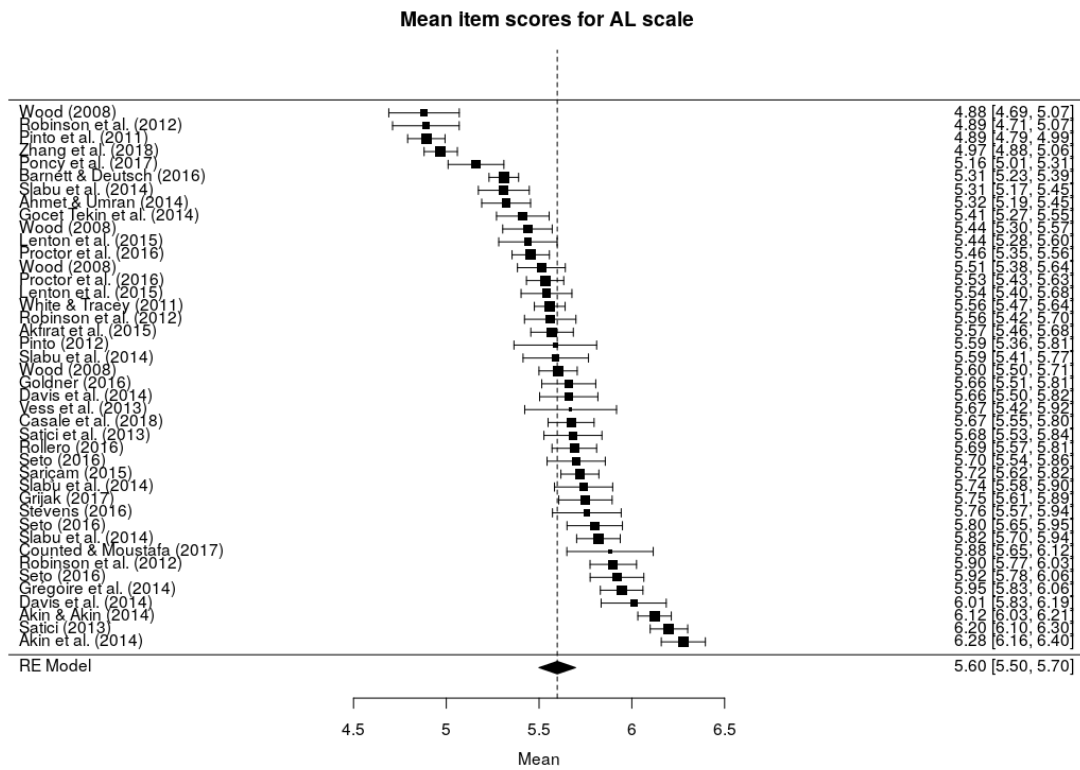


Self-alienation subscale:

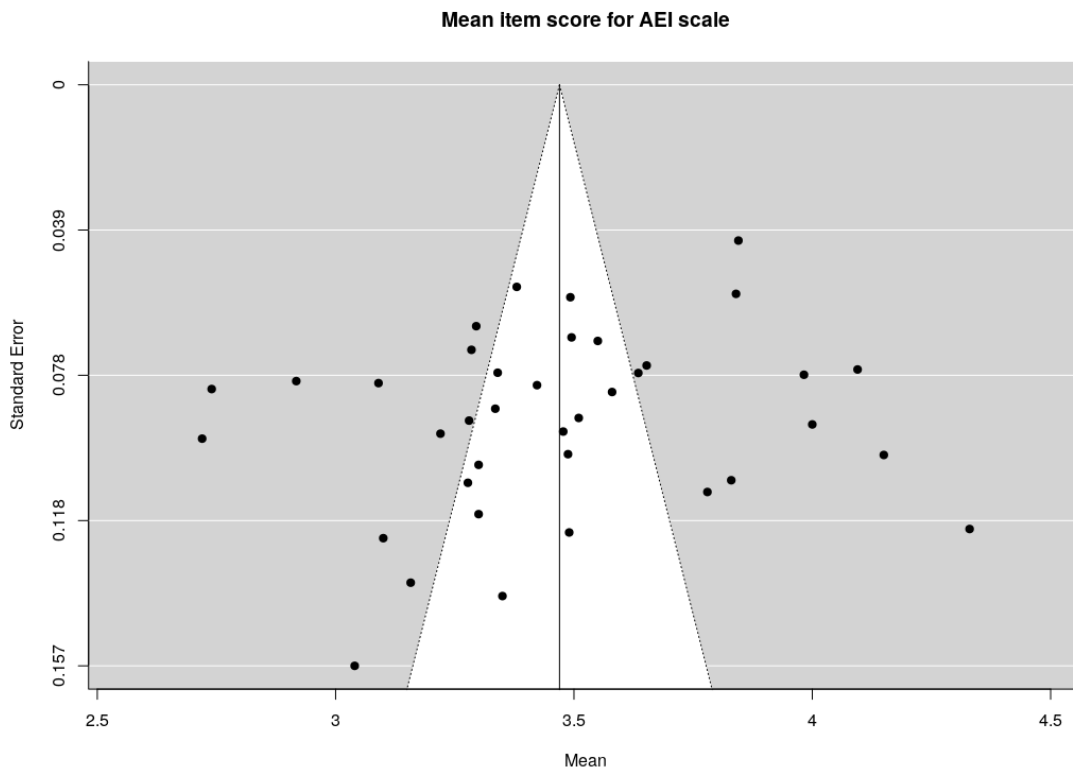
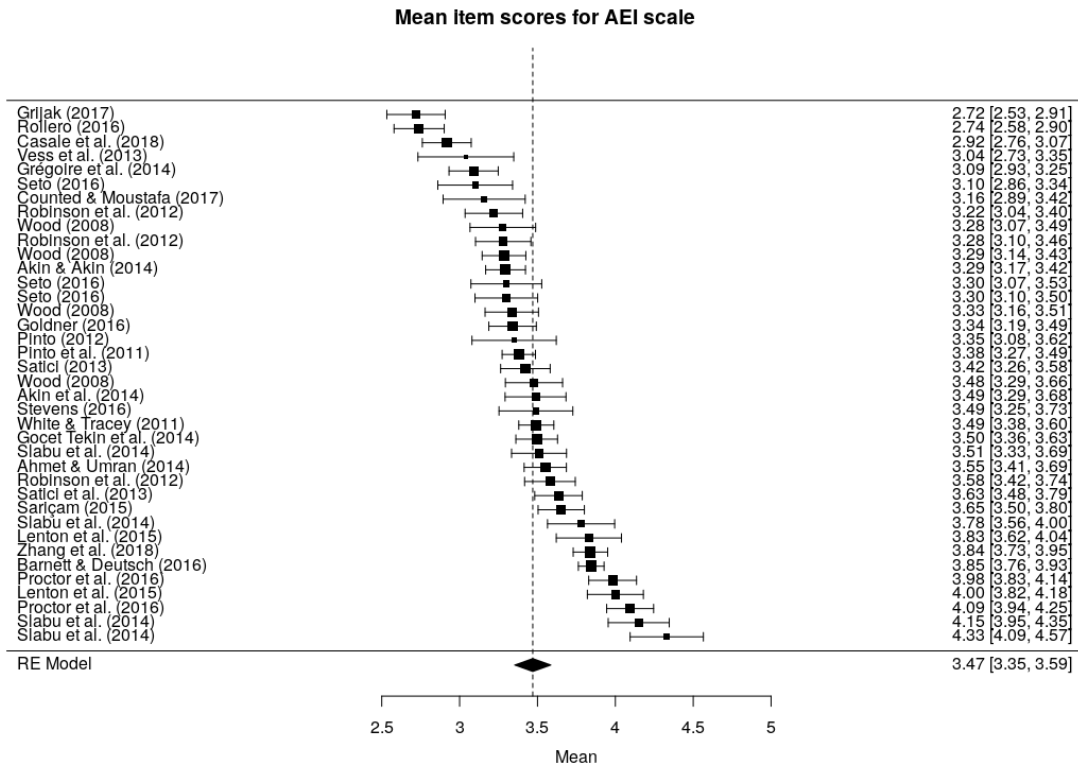


Appendix O: Forest and funnel plots of the mean scores for the AS subscales

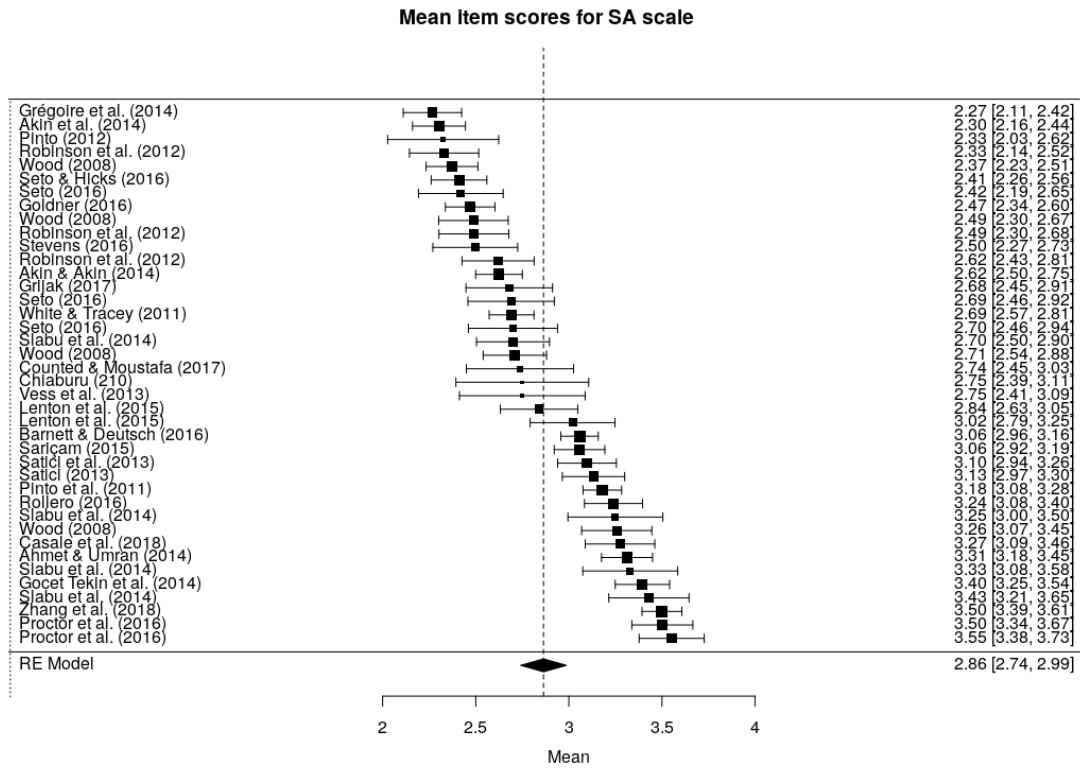
Authentic living subscale:



Accepting external influences subscale:



Self-alienation subscale:

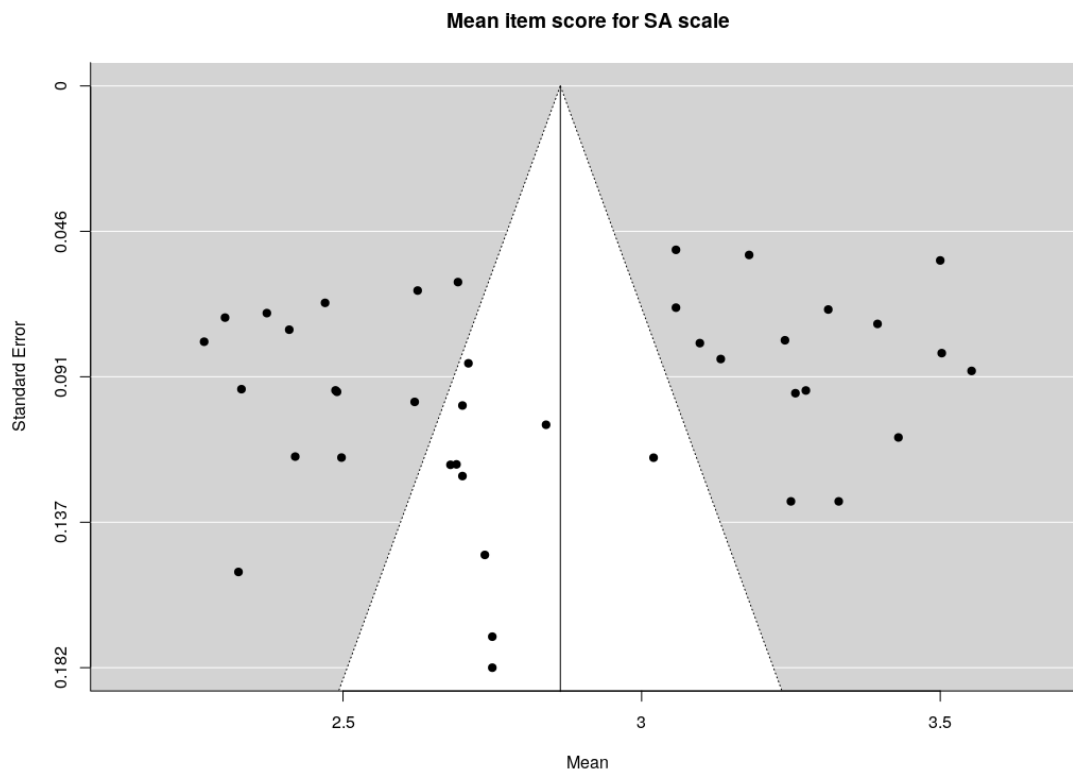


Appendix P: Study 2 - Effects of client demographics

Gender

Baseline. An independent samples t-test revealed baseline AS scores did not differ significantly between males ($M = 47.9, SD = 11.3$) and females ($M = 48.7, SD = 10.7$): $t(56) = 0.22, p = .83$. The mean difference was .72, 95% CI [-5.8, 7.2]. Baseline CORE-10 scores did not differ significantly between males ($M = 18.9, SD = 3.9$) and females ($M = 15.8, SD = 8.4$): $t(25) = -0.99, p = .33$. The mean difference was -3.1, 95% CI [-9.5, 3.3].

Post-therapy. An independent samples t-test revealed post-therapy AS scores did not differ significantly between males ($M = 56.1, SD = 11.8$) and females ($M = 54.7, SD = 12.4$): $t(42) = -0.34, p = .72$. The mean difference was -1.4, 95% CI [-9.8, 7.0]. Post-therapy CORE-10 scores did



not differ significantly between males ($M = 13.0$, $SD = 8.3$) and females ($M = 10.1$, $SD = 8.1$): $t(18) = -0.76$, $p = .46$. The mean difference was -2.9 , 95% CI $[-11.0, 5.1]$.

Change. An independent samples t-test revealed AS change scores did not differ significantly between males ($M = 7.2$, $SD = 15.3$) and females ($M = 6.7$, $SD = 9.3$): $t(38) = -0.12$, $p = .90$. The mean difference was -0.49 , 95% CI $[-8.5, 7.5]$. CORE-10 change scores did not differ significantly between males ($M = -6.1$, $SD = 8.1$) and females ($M = -6.2$, $SD = 8.6$): $t(18) = -0.003$, $p = .99$. The mean difference was -0.01 , 95% CI $[-8.3, 8.3]$.

Age

Baseline. At baseline, age was not significantly correlated with AS scores: $r(45) = -.11$, $p = .23$, 95% CI $[-.38, .18]$. Nor was age significantly correlated with baseline CORE-10 scores: $r(22) = -.18$, $p = .20$, 95% CI $[-.54, .24]$.

Post-therapy. At post-therapy, age was not significantly correlated with AS scores: $r(36) = -.12$, $p = .25$, 95% CI $[-.42, .20]$. Nor was age significantly correlated with post-therapy CORE-10 scores: $r(15) = -.08$, $p = .38$, 95% CI $[-.54, .42]$.

Change. Age did not significantly correlate with either change on the AS, $r(32) = -.15$, $p = .20$, 95% CI $[-.46, .20]$; or the CORE-10, $r(15) = .02$, $p = .47$, 95% CI $[-.47, .50]$.

Ethnicity

Baseline. A one-way ANOVA revealed there was a non-significant effect of ethnicity on baseline AS scores: $F(2, 48) = 1.55, p = .22, \eta_p^2 = .061$. There was also a non-significant effect of ethnicity on baseline CORE-10 scores: $F(2, 21) = 0.24, p = .79, \eta_p^2 = .022$.

Post-therapy. A one-way ANOVA revealed there was a non-significant effect of ethnicity on post-therapy AS scores: $F(2, 36) = 0.88, p = .42, \eta_p^2 = .047$. There was also a non-significant effect of ethnicity on post-therapy CORE-10 scores: $F(2, 14) = 0.67, p = .53, \eta_p^2 = .087$.

Change. A one-way ANOVA revealed there was a non-significant effect of ethnicity on AS change scores: $F(2, 33) = 1.09, p = .35, \eta_p^2 = .062$. There was also a non-significant effect of ethnicity on CORE-10 change scores: $F(2, 14) = 0.03, p = .97, \eta_p^2 = .004$.

Education

Baseline. At baseline, education level shared a significant, moderate positive correlation with AS scores: $rs(22) = .47, p = .010, 95\% \text{ CI } [.08, .73]$. Education was non-significantly correlated with the AL ($rs[22] = .15, p = .25, 95\% \text{ CI } [-.27, .52]$.) and SA ($rs[22] = -.24, p = .13, 95\% \text{ CI } [-.59, .18]$.) subscales at baseline, but shared a significant, large negative correlation with the AEI subscale, $rs(22) = -.62, p = .001, 95\% \text{ CI } [-.82, -.29]$. Participants who were more educated scored higher on the AS overall, and lower on the AEI subscale.

Education level was non-significantly correlated with CORE-10 scores at baseline: $rs(22) = .33, p = .06, 95\% \text{ CI } [-.65, .08]$.

Post-therapy. At post-therapy, education level shared a significant, large positive correlation with AS scores: $r_s(18) = .58, p = .004, 95\% \text{ CI } [.19, .81]$. Education was non-significantly correlated with the AL subscale post-therapy, $r_s(19) = .31, p = .083, 95\% \text{ CI } [-.14, .65]$, but shared significant, large negative correlations with the AEI subscale, $r(19) = -.63, p = .001, 95\% \text{ CI } [-.83, -.27]$, and the SA subscale, $r_s(18) = -.54, p = .007, 95\% \text{ CI } [-.79, .13]$. Post-therapy, more educated participants scored higher on the AS overall, and lower on the AEI and SA subscales.

At post-therapy, education level shared a significant, large negative correlation with CORE-10 scores: $r_s(15) = -.46, p = .032, 95\% \text{ CI } [-.77, .03]$. More educated participants scored lower on the CORE-10 post-therapy.

Change. Education level did not share significant correlations with change on either the AS, $r_s(18) = .15, p = .27, 95\% \text{ CI } [-.31, .56]$, or the CORE-10, $r_s(15) = -.14, p = .30, 95\% \text{ CI } [-.58, .37]$.

Appendix Q: Study 2 - Effects of therapist demographics

Gender

An independent samples t-test revealed client AS change scores did not differ significantly between male therapists ($M = 9.4, SD = 13.5$) and female therapists ($M = 4.6, SD = 9.0$): $t(34) = -1.26, p = .22$. The mean difference was $-4.7, 95\% \text{ CI } [-12.3, 2.9]$. CORE-10 change scores did not differ significantly between male therapists ($M = -9.6, SD = 6.0$) and female therapists ($M = -4.1, SD = 13.5$): $t(15) = 1.38, p = .19$. The mean difference was $5.5, 95\% \text{ CI } [-3.0, 13.9]$.

Age

Therapist age shared a significant, large negative correlation with CORE-10 change, $r(15) = -.57, p = .008, 95\% \text{ CI } [-.82, -.12]$. Clients of older therapists changed more on the CORE-10 over the course of therapy. Therapist age did not significantly correlate with AS change, $r(33) = .16, p = .17, 95\% \text{ CI } [-.18, .47]$.

Ethnicity

A one-way ANOVA revealed the effect of therapist ethnicity on AS change was non-significant, $F(2, 32) = 0.04, p = .96, \eta_p^2 = .003$. There was insufficient data to test the effect of therapist ethnicity on CORE-10 change.

Qualification

Number of years therapists were qualified did not correlate significantly with either AS change ($r[32] = .02, p = .45, 95\% \text{ CI } [-.32, .36]$) or CORE-10 change ($r[13] = -.19, p = .25, 95\% \text{ CI } [-.64, .36]$).

Therapist level of training did not correlate significantly with either AS change ($rs[32] = -.14, p = .22, 95\% \text{ CI } [-.21, .46]$) or CORE-10 change ($rs[13] = .37, p = .09, 95\% \text{ CI } [-.18, .74]$).

Professional identity

A one-way ANOVA revealed the effect of therapist professional identity on AS change was non-significant, $F(1, 32) = 0.55, p = .46, \eta_p^2 = .017$. The effect of therapist professional identity on CORE-10 change was also non-significant, $F(1, 13) = 0.11, p = .75, \eta_p^2 = .008$.

An independent samples t-test revealed that there was no significant difference in AS scores whether the therapist was currently a trainee ($M = 8.0, SD = 9.5$) or not ($M = 2.2, SD = 6.3$): $t(17) = -1.26$. The mean difference was -5.8 , 95% CI $[-15.5, 3.9]$. There was insufficient data to test the effect of therapist trainee status on CORE-10 change.

Therapeutic primary orientation

A one-way ANOVA revealed the effect of therapist primary therapeutic orientation on AS change was non-significant, $F(5, 27) = 0.68, p = .64, \eta_p^2 = .112$. The effect of therapist orientation on CORE-10 change was also non-significant, $F(1, 13) = 0.03, p = .87, \eta_p^2 = .002$.