

**Development of the Brief Addiction Therapist Scale (BATS): A Tool for  
Evaluating Therapist Delivery of Psychological Therapies for Alcohol  
and Drug Use Problems**

Helen Fiona Crosby

Submitted in accordance with the requirements for the degree of  
Doctor of Philosophy

The University of Leeds  
School of Medicine  
Leeds Institute of Health Sciences

September, 2018

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Helen Fiona Crosby to be identified as Author of this work has been asserted by her in accordance with the Copyright, Designs and Patents Act 1988.

© 2018 The University of Leeds and Helen Fiona Crosby

## Acknowledgements

I wish to thank the following people for their considerable generosity. For thought provoking supervision, and much more, I am extremely grateful to my supervisors. My main supervisor, Dr. Gary Latchford, for his unwavering optimism and guidance, not to mention his unsurpassed knowledge of psychotherapy. The good advice, support, and friendship of my second supervisor, Dr. Bridgette Bewick, on both an academic and a personal level has been invaluable. And last, but not least, my mentor, Dr. Gillian Tober, who encouraged me to undertake this research and has supported me over many years.

Thank you to my PhD colleagues, Firdaus Hafidz, Zane Linde-Ozola, and Pritaporn Kingkaew, for their encouragement and many cups of tea; it would have been a lonely experience without you! I am grateful to Dr. Tracey Farragher for her advice on statistics, Ben Grigor for his help with all things IT, and my tutor, Dr. Cathy Brennan, for her support.

Special thanks to my family and friends for their support and encouragement. My partner, Greg, for always being there for me, through the ups and downs, and for his helpful comments on research methodology. My parents who have been so very helpful and who continue to have faith in me. The brilliant conversations with my sister, Janet, that have made me laugh and kept me going these past few years. Beth, Laura, Fiona, Liz, and my softball friends, thank you for your wonderful support and keeping me in touch with reality.

I would like to extend my gratitude to the people who gave their time to take part in the research for this thesis. While individually they must remain anonymous, collectively they have my grateful thanks. Thank you also to the research teams who gave me access to the trial databases and session recordings. For funding this project, I would like to thank Alcohol Research UK and the Society for the Study of Addiction. Many thanks to Magpie for developing the BATS logo and contributing to the scale design.

Without you all, this project would have been an impossible task. Thank you!

## **Abstract**

This thesis describes the development of the BATS: a brief, evidence-based tool monitoring and evaluating therapists' delivery of psychological therapies used in routine practice for alcohol and drug use problems. The BATS is transtheoretical, applicable to the range of widely used therapies in addiction. Four studies were undertaken to develop the BATS. Study 1 identified twenty-six fidelity measures from the literature that evaluate therapists' delivery of psychological therapies for alcohol and drug use problems. Study 2 generated items and response formats for potential inclusion in the BATS using the identified measures as a basis. Generation of the items was primarily based on the results of a thematic analysis; eighteen exemplar items were developed. Study 3 generated a consensus among experts in the fields of addiction and psychotherapy on the content of the BATS. A consensus was obtained using a three-round Delphi survey. At the conclusion of the third round, group agreement on the 12 scale items and response format was reached. This content formed the first version of the BATS. Study 4 tested the psychometric properties of the newly developed scale. The results provided support for inter-rater reliability and convergent validity. The BATS provides a reliable and valid method for evaluating treatment delivery in routine practice, helping to improve our understanding of the process of therapy in addiction. The real world application of the BATS provides a useful tool for training and supervision, which has the potential to impact on therapist competence and treatment delivery.

## Table of Contents

<b>Acknowledgements</b> .....	<b>iii</b>
<b>Abstract</b> .....	<b>iv</b>
<b>Table of Contents</b> .....	<b>v</b>
<b>List of Tables</b> .....	<b>xii</b>
<b>List of Figures</b> .....	<b>xiv</b>
<b>List of Abbreviations</b> .....	<b>xv</b>
List of Abbreviations of the Identified Measures .....	xviii
<b>Chapter 1 Background and aims</b> .....	<b>1</b>
1.1 Introduction .....	1
1.2 Background .....	1
1.3 Aims and objectives.....	5
1.4 Conclusion.....	6
<b>Chapter 2 Research overview</b> .....	<b>7</b>
2.1 Introduction .....	7
2.2 Research methodology .....	7
2.3 Research methods.....	9
2.3.1 Study 1: Identifying existing measures.....	9
2.3.2 Study 2: Generating an item pool .....	9
2.3.3 Study 3: Agreeing the content .....	9
2.3.4 Study 4: Testing reliability and validity.....	10
2.4 Ethical considerations.....	10
2.4.1 Approvals obtained .....	11
2.4.2 Protocol changes.....	11
2.4.2.1 Study 3: Agreeing the scale content .....	11
2.4.2.2 Study 4: Testing reliability and validity .....	11
2.4.3 Summary of the main issues.....	12
2.4.3.1 Consent .....	12
2.4.3.2 Confidentiality .....	13
2.4.3.3 Data storage .....	13
2.4.3.4 Anonymity of responses.....	13
2.5 Conclusion.....	14

<b>Chapter 3 Identifying existing measures .....</b>	<b>15</b>
3.1 Introduction .....	15
3.2 Method .....	15
3.2.1 Selection criteria .....	15
3.2.1.1 Type of articles .....	16
3.2.1.2 Type of measures .....	17
3.2.2 Search methods .....	18
3.2.2.1 Stage 1: Identifying articles .....	18
3.2.2.2 Stage 2: Identifying measures .....	18
3.2.3 Summarising the literature .....	19
3.2.3.1 Characteristics of the identified measures .....	19
3.2.3.2 Methods of validating the identified measures .....	19
3.3 Results .....	19
3.3.1 Outcome of searches .....	19
3.3.1.1 Stage 1: Identifying articles .....	19
3.3.1.2 Stage 2: Identifying measures .....	21
3.3.2 Characteristics of the identified measures .....	21
3.3.2.1 Clinical focus .....	21
3.3.2.2 Purpose .....	22
3.3.2.3 Target therapies .....	22
3.3.2.4 Assessment of fidelity .....	28
3.3.2.5 Number of items .....	30
3.3.2.6 Item scoring .....	30
3.3.3 Methods of validating the identified measures .....	32
3.3.3.1 Validity .....	32
3.3.3.2 Reliability .....	39
3.4 Discussion .....	41
3.4.1 Study overview .....	41
3.4.2 Main findings .....	42
3.4.2.1 Characteristics of the identified measures .....	42
3.4.2.2 Methods of validating the identified measures .....	43
3.4.3 Strengths and limitations .....	45
3.5 Conclusion .....	46
<b>Chapter 4 Generating an item pool .....</b>	<b>47</b>
4.1 Introduction .....	47

4.2 Method .....	47
4.3 Data .....	47
4.4 Data analysis .....	48
4.4.1 Stage 1: Grouping items that ask similar questions .....	48
4.4.1.1 Generating codes .....	49
4.4.1.2 Searching for themes .....	50
4.4.1.3 Reviewing and refining themes .....	50
4.4.1.4 Defining and labelling themes .....	50
4.4.2 Stage 2: Selecting themes most relevant to the BATS .....	51
4.4.3 Stage 3: Generating items for the BATS .....	51
4.4.3 Stage 4: Generating response formats for the BATS.....	52
4.5 Results.....	52
4.5.1 Stage 1: Grouping items that ask similar questions .....	52
4.5.1.1 Session management .....	54
4.5.1.2 Medication and case management.....	59
4.5.1.3 Interventions to increase awareness .....	61
4.5.1.4 Interventions to change behaviour .....	71
4.5.1.5 Core skills.....	78
4.5.2 Stage 2: Selecting themes for the BATS .....	87
4.5.3 Stage 3: Generating items for the BATS .....	91
4.5.4 Stage 4: Generating response formats for the BATS.....	91
4.6 Discussion .....	94
4.6.1 Study overview.....	94
4.6.2 Main findings .....	94
4.6.2.1 Items for the BATS .....	94
4.6.2.2 Response formats for the BATS .....	95
4.6.3 Strengths and limitations.....	95
4.7 Conclusion.....	96
<b>Chapter 5 Agreeing the content .....</b>	<b>97</b>
5.1 Introduction .....	97
5.2 Method .....	97
5.2.1 Participant selection.....	97
5.2.2 Measures .....	99
5.2.2.1 Demographic questionnaire .....	99
5.2.2.2 Round one questionnaire.....	99

5.2.2.3 Round two questionnaire .....	100
5.2.2.4 Round three questionnaire .....	101
5.2.3 Procedure .....	102
5.2.4 Data analysis.....	103
5.2.4.1 Participant data .....	103
5.2.4.2 Item data .....	103
5.2.4.3 Item scoring data .....	104
5.2.4.4 Consensus .....	105
5.2.4.5 General feedback.....	106
5.3 Results .....	106
5.3.1 Participants.....	106
5.3.2 Round one .....	108
5.3.2.1 Item data .....	108
5.3.2.2 General feedback.....	113
5.3.3 Round two .....	114
5.3.3.1 Item data .....	114
5.3.3.2 Item scoring data .....	117
5.3.3.3 General feedback.....	118
5.3.4 Round three.....	118
5.4.3.1 Item data .....	118
5.4.3.2 General feedback.....	123
5.4 Discussion .....	124
5.4.1 Study overview .....	124
5.4.2 Main findings .....	125
5.4.2.1 Item data .....	125
5.4.2.2 Item scoring data .....	126
5.4.2.3 General feedback.....	126
5.4.3 Study strengths and limitations.....	126
5.4.3.1 Participants .....	126
5.4.3.2 Questionnaires .....	128
5.4.3.3 Analyses .....	128
5.5 Conclusion .....	129
<b>Chapter 6 Testing reliability and validity .....</b>	<b>130</b>
6.1 Introduction .....	130
6.2 Method .....	130



6.2.1	Data sources .....	131
6.2.2	Secondary analysis of trial data .....	131
6.2.2.1	ADAPTA .....	132
6.2.2.2	AESOPS .....	133
6.2.2.3	UKATT .....	134
6.2.3	Routine practice data .....	135
6.2.3.1	NHS specialist addictions service (SAS) .....	135
6.2.3.2	Non-NHS drug and alcohol service (DAS) .....	136
6.2.4	Process rating data derived from the BATS .....	137
6.2.4.1	Selecting recordings for process rating .....	138
6.2.4.2	The process of rating .....	140
6.2.5	Data analysis .....	140
6.2.5.1	Sessions rated using the BATS .....	140
6.2.5.2	Item scores on the BATS .....	141
6.2.5.3	Convergent validity .....	141
6.2.5.4	Inter-rater reliability .....	144
6.3	Results .....	146
6.3.1	Sessions rated using the BATS .....	146
6.3.2	Item scores on the BATS .....	147
6.3.3	Convergent validity .....	149
6.3.3.1	Relationship to the Working Alliance Inventory .....	149
6.3.3.2	Relationship to the process rating measures .....	151
6.3.4	Inter-rater reliability .....	152
6.4	Discussion .....	152
6.4.1	Study overview .....	152
6.4.2	Main findings .....	154
6.4.2.1	Sessions rated using the BATS .....	154
6.4.2.2	Item scores on the BATS .....	155
6.4.2.3	Convergent validity .....	155
6.4.2.4	Inter-rater reliability .....	156
6.4.3	Study strengths and limitations .....	157
6.4.3.1	Sample size .....	157
6.4.3.2	Weighted kappa versus the ICC .....	158
6.5	Conclusion .....	159

<b>Chapter 7 General discussion .....</b>	<b>161</b>
7.1 Introduction.....	161
7.2 Project overview .....	161
7.3 Summary of the findings .....	162
7.3.1 Study 1: Identifying existing measures .....	162
7.3.2 Study 2: Generating an item pool.....	163
7.3.2.1 Items for the BATS.....	163
7.3.2.2 Response formats for the BATS.....	164
7.3.3 Study 3: Agreeing the content.....	164
7.3.3.1 Items for the BATS.....	164
7.3.3.2 Response formats for the BATS.....	165
7.3.4 Study 4: Testing reliability and validity .....	166
7.4 The BATS within the context of psychotherapy research .....	167
7.5 Strengths and limitations .....	170
7.6 Clinical implications and future directions .....	173
7.7 Conclusion .....	174
<b>References .....</b>	<b>175</b>
<b>Appendix A Chapter 2 Appendices .....</b>	<b>192</b>
A.1 Regulatory approvals – NRES committee.....	193
A.2 Regulatory approvals – NHS permission .....	198
A.3 Regulatory approvals – University ethics committee .....	202
A.4 Regulatory approvals – Amendment 1 .....	204
<b>Appendix B Chapter 3 Appendices .....</b>	<b>208</b>
B.1 Search strategy for identifying existing fidelity measures .....	209
B.2 Summary of fidelity measures and associated development and/or validation articles excluded from the literature review .....	212
B.3 Summary of therapies covered by the identified fidelity measures.....	220
B.4 Summary of the development and validation of the identified measures .....	224
B.5 Methods for establishing construct validity .....	238
B.6 Variants of the intraclass correlation coefficient.....	239
<b>Appendix C Chapter 4 Appendices .....</b>	<b>240</b>
C.1 Refinements to the thematic structure made by the researcher and the supervision team.....	241
<b>Appendix D Chapter 5 Appendices .....</b>	<b>248</b>
D.1 Participant information screenshot from round one .....	249

D.2	Consent page screenshot from round one .....	252
D.3	Example screenshot from the first round questionnaire .....	253
D.4	Participants grouped by demographic information to gauge individual participation across the rounds.....	255
D.5	Statistical summary of participants' ratings in round one .....	256
D.6	Statistical summary of participants' ratings in round two .....	257
D.7	The BATS.....	258
	<b>Appendix E Chapter 6 Appendices .....</b>	<b>260</b>
E.1	Routine practice recordings therapist information sheet .....	261
E.2	Routine practice recordings therapist consent form.....	265
E.3	Routine practice recordings therapist questionnaire.....	266
E.4	Routine practice recordings client covering letter .....	267
E.5	Routine practice recordings client information sheet.....	268
E.6	Routine practice recordings client consent form .....	271
E.7	Routine practice recordings friends and family information sheet.....	272
E.8	Routine practice recordings client questionnaire .....	275
E.9	Ordinary least squares and weighted least squares methods .....	276
E.10	Matching WAI data with corresponding sessions in the BATS dataset .....	278
E.11	Calculating Kendall's tau-b for the BATS and the WAI-S client. ....	279
E.12	Selected items from the trial process rating measures for the convergent validity analyses.....	281
E.13	Checks of normality for data used to examine convergent validity .....	284
E.14	Calculating weighted kappa for item 1 'problem focused' .....	285
E.15	Calculating intraclass correlation coefficient for item 1 'problem focused' .....	288
E.16	Sample size estimation for the ICC.....	291

## List of Tables

Table 1: An overview of the participants recruited .....	13
Table 2: Criteria for article and measure selection .....	16
Table 3: Characteristics of the identified fidelity measures.....	23
Table 4: Response formats used in the AESOPS PRS .....	31
Table 5: Summary the main tests of reliability and validity .....	44
Table 6: Phases of the thematic analysis .....	49
Table 7: An example of how items were coded using multiple and exclusive codes .....	50
Table 8: Examples of how themes were refined.....	51
Table 9: An overview of ‘session management’ (meta-theme 1) with theme descriptions and exemplar items.....	57
Table 10: An overview of ‘medication and case management’ (meta-theme 2) with theme descriptions and exemplar items.....	60
Table 11: An overview of ‘interventions to increase awareness’ (meta-theme 3) with theme descriptions and exemplar items.....	68
Table 12: An overview of ‘interventions to change behaviour’ (meta-theme 4) with theme descriptions and exemplar items.....	76
Table 13: An overview of ‘core skills’ (meta-theme 5) with theme descriptions and exemplar items.....	84
Table 14: Summary of the revised thematic structure .....	88
Table 15: Generated items for potential inclusion in the BATS .....	92
Table 16: Participation across rounds .....	106
Table 17: Participants’ self-reported participation across rounds .....	107
Table 18: Participant characteristics across all three rounds.....	108
Table 19: Median and interquartile range (IQR) for participants’ ratings of item importance and comprehensibility in round one.....	110
Table 20: Participants’ comments, and ratings of importance and comprehensibility for item 1 ‘problem focused’ in round one.....	112
Table 21: Items removed in round one .....	113
Table 22: Median and interquartile range (IQR) for participants’ ratings of item importance and comprehensibility in round two.....	115
Table 23: Items removed in round two .....	117
Table 24: Percentage agreements, median, interquartile range (IQR), and range for participants’ ratings of item importance in round three.....	119

<b>Table 25: Statistical summaries of participants' ratings for importance and comprehensibility across rounds .....</b>	<b>120</b>
<b>Table 26: Item-specific feedback for items in round three .....</b>	<b>122</b>
<b>Table 27: Number of items retained, revised and removed across rounds.....</b>	<b>125</b>
<b>Table 28: Data sources used in the development of the BATS .....</b>	<b>131</b>
<b>Table 29: Coding design for the seven-category 'therapy type' variable .....</b>	<b>141</b>
<b>Table 30: Assessing associations with Kendall's tau-b .....</b>	<b>143</b>
<b>Table 31: Ratings weighted with quadratic weights .....</b>	<b>145</b>
<b>Table 32: Assessing agreement with weighted kappa .....</b>	<b>145</b>
<b>Table 33: Assessing agreement with ICC.....</b>	<b>146</b>
<b>Table 34: Summary of the length of the sessions .....</b>	<b>147</b>
<b>Table 35: Differences in duration between routine practice and trial therapies .....</b>	<b>147</b>
<b>Table 36: Summary of the item scores across therapies .....</b>	<b>148</b>
<b>Table 37: Differences in total scores between routine practice and trial therapies.....</b>	<b>149</b>
<b>Table 38: Number of cases in the BATS dataset with corresponding WAI data .....</b>	<b>149</b>
<b>Table 39: Relationships between the BATS and the WAI-S client (n=24) .....</b>	<b>150</b>
<b>Table 40: Relationships between the BATS and the WAI-S therapist (n=31).....</b>	<b>150</b>
<b>Table 41: Relationships between the trial-specific fidelity measures and the BATS (n=20).....</b>	<b>151</b>
<b>Table 42: Weighted kappa and intraclass correlation coefficients for items on the BATS.....</b>	<b>153</b>

## List of Figures

Figure 1: Flowchart illustrating the selection of measures included in the review .....	20
Figure 2: Relationships between therapies covered by the identified measures .....	29
Figure 3: Methods for generating items and response formats for the BATS .....	48
Figure 4: Thematic map showing the five meta-themes and associated themes .....	53
Figure 5: Response formats generated for the BATS .....	91
Figure 6: Screenshot from the first round questionnaire of the 7-point scales used to rate item importance and comprehensibility .....	100
Figure 7: Screenshot from the second round questionnaire of the proposed response formats for the BATS .....	101
Figure 8: Screenshot from the third round questionnaire of the summary of results from round two for item 1 .....	102
Figure 9: Frequency distributions of participants' importance and comprehensibility ratings for item 1 'problem focused' in round one .....	109
Figure 10: Frequency distributions of participants' importance and comprehensibility ratings for item 1 'problem focused' in round two .....	116
Figure 11: Frequency distributions of participants' importance ratings for item 1 'problem focused' in round three .....	123
Figure 12: Study 3 overview .....	124
Figure 13: Flow diagram of the selection of recordings for independent and double rating .....	139

## List of Abbreviations

ACS-IDCCD	Adherence/Competence Scale for Individual Drug Counseling for Cocaine Dependence
ACS-SEC	Adherence/Competence Scale for Supportive-Expressive Therapy for Cocaine Dependence
ADAPTA	Addressing Drinking Among Patients: comparing Two Approaches
ADAPTA PRS	ADAPTA (Addressing Drinking Among Patients: comparing Two Approaches) Process Rating Scale
AESOPS PRS	AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale
AF	Alcohol focused intervention
ANOVA	Analysis of variance
ATRS	Addiction Therapist Rating Scale
BA	Brief advice
BAS	Brief Negotiation Interview Adherence Scale
BATS	Brief Addiction Therapist Scale
BECCI	Behaviour Change Counselling Index
BCC	Behaviour change counselling
BNI	Brief Negotiation Interview
CB	Cognitive-behavioural
CBT	Cognitive-behavioural therapy
CBT for PTSD Fidelity Scale	Cognitive-Behavioral Therapy for Post-Traumatic Stress Disorder Fidelity Scale
CBT Therapist Checklist	Cognitive Behavioral Therapy Therapist Checklist
CE	Compliance enhancement
CE Therapist Rating Form	Compliance Enhancement Therapist Adherence/Competence Rating Form
CFA	Confirmatory factor analysis
CI	Confidence interval
CLAMI	Client Language Assessment in Motivational Interviewing
ClinM	Clinical management
CM	Contingency management
CM Rating Form	Contingency Management Clinician Adherence/Competence Rating Form

CPPS	Comparative Psychotherapy Process Scale
CRN	Clinical Research Network
CSPRS-6	Collaborative Study Psychotherapy Rating Scale – Form 6
CT	Cognitive therapy
CTACS	Cognitive Therapy Adherence and Competence Scale
CTS	Cognitive Therapy Scale
CTS-R	Cognitive Therapy Scale – Revised
CVI	Content validity index
DAS	Drug and alcohol service (non-NHS)
DVD	Digital video disc
EFA	Exploratory factor analysis
ET	Exploratory therapy
FFM	Friend or family member
GROMIT	Global Rating of Motivational Interviewing Therapist
HL	Healthy living intervention
LIHS	Leeds Institute of Health Sciences
IAC	I Am Concerned
IAC Treatment Fidelity Instrument	IAC (I Am Concerned) Treatment Fidelity Instrument
IBM	International Business Machines
ICC	Intraclass correlation coefficient
IDC	Individual drug counselling
IQR	Interquartile range
IPT	Interpersonal therapy
IRR	Inter-Rater Reliability
ITRS	Independent Tape Rater Scale
MACT	Manual Assisted Cognitive Therapy
MACT Rating Scale	Manual Assisted Cognitive Therapy Rating Scale
MET	Motivational enhancement therapy
MI	Motivational interviewing
MI-CBT	Integrated motivational interviewing and cognitive-behavioural therapy
MI-CTS	Integrated Motivational Interviewing and Cognitive Behavioural Therapy Fidelity Scale
MISTS	Motivational Interviewing Supervision and Training Scale
MISTS Revised	Motivational Interviewing Supervision and Training Scale Revised



MITI	Motivational Interviewing Treatment Integrity Code
MTRS	Minnesota Therapy Rating Scale
NHS	National Health Service
NRES	National Research Ethics Service
OLS	Ordinary least squares
PCA	Principle components analysis
PI	Psychodynamic-interpersonal
PIS	Participant information sheet
PTSD	Post-traumatic stress disorder
RCT	Randomised controlled trial
RP	Routine practice
SAS	Specialist addictions service (NHS)
SBNT	Social behaviour and network therapy
SD	Standard deviation
SET	Supportive-expressive therapy
SPR	Systems for Psychotherapy Ratings
SPRS	Sheffield Psychotherapy Rating Scale
SPSS	Statistical Package for Social Sciences
SRM	Standardised response mean
TA	Thematic analysis
TPRS	Therapy Process Rating Scale
TSF	Twelve-step facilitation
TSF-ACES	Twelve-Step Facilitation Adherence Competence Empathy Scale
UK	United Kingdom
UKATT	United Kingdom Alcohol Treatment Trial
UKATT PRS	UKATT (United Kingdom Alcohol Treatment Trial) Process Rating Scale
WAI	Working Alliance Inventory
WAI-S	Working Alliance Inventory – Short
WLS	Weighted least squares
YACS	Yale Adherence and Competence Scale
YACSII	Yale Adherence and Competence Scale Second Edition

## List of Abbreviations of the Identified Measures

ACS-SEC	Adherence/Competence Scale for Supportive-Expressive Therapy for Cocaine Dependence
ACS-IDCCD	Adherence/Competence Scale for Individual Drug Counseling for Cocaine Dependence
ADAPTA PRS	ADAPTA (an alcohol-focused intervention versus a healthy living intervention for problem drinkers identified in a general hospital setting) Process Rating Scale
AESOPS PRS	AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale
BAS	Brief Negotiation Interview Adherence Scale
BECCI	Behaviour Change Counselling Index
CBT for PTSD Fidelity Scale	Cognitive-Behavioral Therapy for Post-Traumatic Stress Disorder
CBT Therapist Checklist	Cognitive Behavioural Therapy Therapist Checklist
CE Therapist Rating Form	Compliance Enhancement Therapist Adherence/Competence Rating Form
CM Clinician Rating Form	Contingency Management Clinician Adherence/Competence Rating Form
CPPS	Comparative Psychotherapy Process Scale
CSPRS-6	Collaborative Study Psychotherapy Rating Scale – Form 6
CTACS	Cognitive Therapy Adherence Scale
CTS-R	Cognitive Therapy Scale – Revised
GROMIT	Global Rating of Motivational Interviewing Therapist
IAC Treatment Fidelity Instrument	I Am Concerned (a brief opportunistic intervention for prenatal substance use) Fidelity Instrument
ITRS	Independent Tape Rater Scale
MACT Rating Scale	Manual Assisted Cognitive Therapy Rating Scale
MI-CTS	Integrated Motivational Interviewing and Cognitive Behavioural Therapy Fidelity Scale
MISTS Revised	Motivational Interviewing Supervision and Training Scale Revised
MTRS	Minnesota Therapy Rating Scale
SPRS	Sheffield Psychotherapy Rating Scale
TPRS	Therapy Process Rating Scale
TSF-ACES	Twelve Step Facilitation Adherence Competence Empathy Scale
UKATT PRS	UKATT (United Kingdom Alcohol Treatment Trial) Process Rating Scale
YACSII	Yale Adherence and Competence Scale Second Edition

# Chapter 1

## Background and aims

### 1.1 Introduction

This chapter introduces the current project. The project aimed to develop the Brief Addiction Therapist Scale (BATS), a tool for evaluating therapist delivery of psychological therapies used for alcohol and drug use problems in routine practice. This chapter introduces the background literature, and explains why the project was conducted. The project comprised four studies, ensuring the BATS was developed as comprehensively as possible. This chapter presents the overarching project aim, as well as the objectives for each of the four studies.

### 1.2 Background

The harms related to alcohol and other drugs are well documented. Alcohol consumption is the world's leading risk factor for disease and disability after smoking and obesity (World Health Organization, 2011). Alcohol has been shown to be a contributory factor in over 60 types of disease and injury, such as, high blood pressure, liver cirrhosis, and various cancers (National Institute for Health and Clinical Excellence, 2011). The number of alcohol-related hospital admissions is increasing. In England, an estimated 337 thousand people were admitted to hospital in 2016/2017 with alcohol misuse problems, an increase of 17% from 2006/2007 (NHS Digital, 2018). There is also a wealth of evidence on the adverse health effects of drug use (Imtiaz et al., 2016); for example, cannabis use is associated with an increased risk of anxiety and depression (Volkow et al., 2014), and opioid dependence and injecting are significant contributors to the global burden of disease (Degenhardt et al., 2013).

The detrimental impact of alcohol and drug use has led to the development of effective psychological therapies. Effective therapies include: motivational enhancement therapy (MET), social behaviour and network therapy (SBNT), and cognitive-behavioural therapy (CBT) (Miller et al., 1992; Copello et al., 2002; Carroll, 1998). The standard for determining the effectiveness of psychological therapies is the randomised controlled trial (RCT)

(Bothwell et al., 2016). One of the reasons why RCTs are considered the 'gold standard' for therapeutic evidence is that they minimise the risk of confounding, any differences in outcome can be attributed to the treatment intervention. The powerful RCT design has replaced earlier scepticism about the ability of psychological therapies to improve the outcomes of clients with alcohol and drug use problems (Longabaugh et al., 2005).

One way to minimise the risk of confounding is to standardise the treatments delivered. As Green and Latchford (2012) explain: "*if therapy were delivered in very different ways by the different therapists in a trial, the results might be hard to interpret*" (p.27). Standardising the way therapies are delivered is achieved with treatment manuals. A treatment manual specifies the underlying principles and techniques of a particular psychological therapy (Heppner et al., 2016); one of the most well-known manuals is the 'Cognitive Therapy of Depression' (Beck et al., 1979). In research trials, treatment manuals are useful for reducing the variability between therapists, and for ensuring that the treatments are correctly delivered (Wampold and Imel, 2015). Manuals are also useful for supporting the implementation of treatments in different clinical settings (Chambless and Hollon, 2012) – treatments can be tested and compared. However, having access to a manual does not guarantee that a therapy will be delivered as intended; to ascertain adherence to manualised therapies, treatment delivery must be assessed for fidelity (Schoenwald et al., 2011).

Treatment fidelity, sometimes termed treatment integrity, has three components. The first component, treatment adherence, refers to the degree to which therapists deliver the specific techniques that characterise a particular therapy as described in the treatment manual (Schoenwald et al., 2011). The second component, therapist competence, refers to the level of skill and judgement shown by therapists in delivering the therapy (Perepletchikova et al., 2007). The third component, treatment differentiation, refers to whether the therapies are distinguishable from one another (Waltz et al., 1993). Most definitions of treatment fidelity emphasise the first two components. This is because treatment adherence and differentiation are closely related: "*a measure of adherence is sufficient to determine whether treatments are different*" (Perepletchikova and Kazdin, 2005, p.366).

A breakdown in either treatment adherence or therapist competence may compromise treatment fidelity and threaten the internal validity of the study (Perepletchikova and

Kazdin, 2005). For example, if therapists deliver techniques proscribed in the treatment manual, inferences about the results obtained would be misleading. If therapists adhered to the manual but delivered the prescribed techniques with limited skill, it may affect treatment progress and contribute to poor outcomes. Treatment fidelity is, therefore, important for evaluating treatment effectiveness as it enables treatment effects to be accurately attributed (Tober et al., 2008).

To assess fidelity requires reliable and valid measures to monitor therapists' delivery of treatments. One of the earliest measures to be developed in the addiction field was the MATCH (Matching Alcohol Treatments to Client Homogeneity) Tape Rating Scale (Carroll et al., 1998a). Adapted from assessment methods used in the National Institute of Mental Health Treatment of Depression Collaborative Research Program (Hill et al., 1992), the measure incorporated Likert-type items to assess adherence and differentiate between treatments. The measure was developed for use in Project MATCH, the largest study of psychological therapies for alcohol use problems (Project MATCH Research Group, 1997). Three treatments were compared: MET, CBT, and twelve-step facilitation (TSF). Process assessments (using the MATCH Tape Rating Scale) showed that the treatments were distinguishable from one another in that the therapists delivered the treatments as described in the treatment manuals (Carroll et al., 1998a). While initial support for reliability and validity of the measure was provided, only one facet of fidelity was evaluated; the measure did not assess therapist competence.

A more comprehensive measure is the Yale Adherence and Competence Scale (YACS). YACS was designed to assess adherence and competence in the delivery of therapies commonly used for substance use problems (Carroll et al., 2000). The measure included three general scales assessing general aspects of substance use treatment (e.g., goals of treatment), and three treatment specific scales measuring the specific ingredients of three treatments: CBT, TSF, and clinical management. The measure has demonstrated validity and reliability, and has specific utility in studies comparing multiple therapies. As the YACS is a research tool, its use in other settings (e.g., training and supervision contexts) is limited (Madson and Campbell, 2006). In contrast, the Motivational Interviewing Supervision Training Scale (MISTS) was specifically developed to monitor the training and supervision of therapists delivering motivational interviewing (Madson et al., 2005). MISTS showed potential for use in training and supervision, but has yet to be evaluated outside of the research environment (Madson and Campbell, 2006).

To be effective in routine practice, treatments shown to be effective in research trials need to be delivered with fidelity. Fidelity monitoring supports the continued effectiveness of evidence based treatments and promotes continuous quality improvement (Aarons et al., 2011). Fidelity to addiction therapies is, in theory, currently monitored during clinical supervision. Clinical supervision is defined as *“the formal provision by senior/qualified health practitioners of an intensive relationship-based education and training that is case-focused and which supports, directs and guides the work of colleagues”* (Milne, 2007, p.440). Supervision provides a forum for therapists to reflect on current practice and to improve competence (Care Quality Commission, 2008).

Use of supervision to assess treatment fidelity currently relies heavily on self-reports. Self-reports increase the potential for inaccuracy due to poor recollection. There is also a tendency for therapists to be overly positive in their evaluations of adherence and competence (Breitenstein et al., 2010; Martino et al., 2009a). As a result, self-reports should not be solely relied on to monitor treatment fidelity in routine practice (Carroll et al., 2010). Video recordings of therapy sessions offer one solution, although objectivity of supervisors’ evaluation of clinical practice cannot be assured. Without an objective means of evaluating treatment delivery the assessment is open to bias and individual variation (Tweed et al., 2010). Treatment fidelity measures are, therefore, needed to support implementation and sustainability of evidence-based practices.

Measures of treatment fidelity have been developed within the context of research trials, but it is argued that there is also a role for such measures in routine practice, not least in informing clinical supervision. For example, supervisors could use a fidelity measure to review a therapy session and provide feedback regarding the therapist’s relative strengths and weaknesses (Bassett et al., 2016). Such a measure would also support peer supervision; if therapists completed a fidelity measure after viewing a therapy session, the measures could be compared, allowing for any *“differences in the interpretation of the session to be discussed”* (Bassett et al., 2016, p.11).

Existing fidelity measures are unsuitable for use in routine practice because they are either too long (e.g., YACS; Carroll et al., 2000) or specific to one therapy (e.g., Madson et al., 2005). In routine practice, therapists use a range of therapies, from brief advice to intensive specialist treatment, to address clients’ alcohol and drug use problems (Raistrick et al., 2006). These therapies are tailored by therapists to meet individual client needs

(Norcross and Wampold, 2011). Therapists will often respond flexibly drawing on techniques from different therapies. To have utility in routine practice, a transtheoretical scale is needed to evaluate therapist delivery of evidence-based therapies for drug and alcohol use problems.

### **1.3 Aims and objectives**

This project aimed to develop the Brief Addiction Therapist Scale (BATS): a tool for evaluating the delivery of psychological therapies widely used for alcohol and drug use problems in routine practice.

The development of the BATS comprised four studies. The objectives of each of the four studies are as follows:

- i) **Study 1 'identifying existing measures'**: The primary objective of Study 1 was to identify existing fidelity measures from the literature that evaluate delivery of psychological therapies used for addressing alcohol and drug use problems. The measures identified from the literature formed the basis of generating items for the BATS in Study 2. The secondary objective was to review the methods used to develop and validate the identified measures; this was important for informing the psychometric work of the BATS in Study 4.
- ii) **Study 2 'generating an item pool'**: The primary objective of Study 2 was to generate items for potential inclusion in the BATS using the fidelity measures identified in Study 1 as a basis. At this stage, consideration was also given to scoring. The secondary objective, therefore, was to generate response formats for the BATS. Generating response formats was important for allowing the items to be written in a consistent format.
- iii) **Study 3 'agreeing the content'**: The objective of Study 3 was to obtain a consensus among selected experts in the fields of addiction and psychotherapy on the content of the BATS using Delphi methodology. This was important for supporting the development of a more robust measure. That is, the expert group were in a position to provide a more updated exchange of information than the researcher and supervision team (GL, BB, and GT) in deciding on the items and response format for the BATS.

- iv) **Study 4 ‘testing reliability and validity’**: The final study aimed to test the psychometric properties of the BATS. Specifically, the study aimed to conduct convergent validity, and inter-rater reliability analyses. These analyses were informed by the findings of Study 2 on the methods used to develop and validate the identified fidelity measures. This psychometric work was important for ensuring the BATS would provide a useful tool for training and supervision.

#### **1.4 Conclusion**

The harms related to alcohol and other drugs are well documented. This has led to the development of effective psychological therapies. The standard for determining the effectiveness of psychotherapies is the RCT (Bothwell et al., 2016). In RCTs, treatment manuals are used to standardise the way therapies should be delivered (Green and Latchford, 2012). However, having access to a manual does not guarantee that the therapies will be delivered as intended; to ascertain adherence to manualised therapies, treatment delivery must be assessed for fidelity (Schoenwald et al., 2011). When evaluating treatment effectiveness, fidelity monitoring is important as it enables treatment effects to be accurately attributed (Tober et al., 2008).

Therapies with a strong evidence base need to be delivered with fidelity in routine practice to support implementation and sustainability of evidence-based practices. In routine practice, therapists use a range of psychological therapies, from brief advice to intensive specialist treatment, to address alcohol and drug use problems (Raistrick et al., 2006). Therapists will often respond flexibly drawing on techniques from different therapies, and tailor interventions to meet client needs (Norcross and Wampold, 2011). To have utility in routine practice, a transtheoretical scale is needed to evaluate the range of widely used therapies for alcohol and drug use problems. This project, therefore, aimed to develop the BATS, a transtheoretical tool for monitoring treatment delivery in routine practice. The next chapter provides an overview of how the BATS was developed.



## Chapter 2

### Research overview

#### 2.1 Introduction

The previous chapter introduced the current project, which set out to develop the BATS. The aims and objectives provided an indication of how the project was conducted. This chapter focuses on the research methodology and the methods used to develop the BATS. The ethical considerations associated with the project are presented, including the approvals obtained for the project, and a summary of the main ethical issues (e.g., consent, and data storage).

#### 2.2 Research methodology

This project adopted a positivist approach<sup>1</sup>. Positivism has a singular view of reality; that there is a physical existence, and that an individual's interaction with that existence constitutes reality. In other words, there is a *"relationship between the world (objects, events, phenomena) and our perception, and understanding, of it"* (Willig, 2013, p.3). Knowledge of the world as measured by the positivist is derived from experience. Comte<sup>2</sup> argued the existence of a distinction between the *"observable"* (i.e., reality) and the *"factitious"* (that which we cannot perceive) (Giddens, 1995, p.148). For positivists, progress is made by means of the experimental method (Delanty, 2005): researchers search for objective, universal laws through the application of carefully defined methods (Patton, 2002). The use of rigorously formatted processes allows for the replication and verification of results. Emphasising the use of rigorous methods allows positivists to ascribe *"certainty of knowledge"*, with outputs of research being considered *"reflections of reality"* (Cruickshank, 2012, p.72). In general, the laws positivists seek to uncover are causal (Delanty, 2005); researchers focus on identifying and examining the factors that influence outcomes (Cresswell, 2009). A positivist approach is dominant within the natural sciences, including health, particularly for quantitative methods of inquiry.

---

<sup>1</sup> This approach refers to the general tradition of positivism, rather than adopting a particular form, such as, logical positivism, or post-positivism.

<sup>2</sup> Auguste Comte (1798-1857), a French philosopher best known for his work on positivism within the field of sociology.

This project is consistent with a positivist position for three reasons. First, the project assumed that there are therapeutic techniques common to therapies widely used in the addiction field. These techniques underpin the development of the BATS, and are themselves the result of a positivistic approach to the scientific method. Thus, coherency with preceding and deceding studies is maintained. Second, the project aimed to develop the BATS by using rigorous and well defined methods that are testable and repeatable. As such, the project aimed to produce a 'positive' outcome that can be objectively viewed by peers. Third, testing of the measure was undertaken in a manner that views the outcome as offering an objective 'truth' as to the psychometric properties of the BATS.

There are two main limitations to using a positivist position. The first limitation is in transposing ontological questions about the existence of reality into epistemological questions about our knowledge or experience of that reality – the “*epistemic fallacy*” (Cruickshank, 2012, p.72). For positivists, the perception of reality is based on what can be observed. This truncated view of reality is problematic, as it ignores the existence of unobserved causal structures. Because the BATS was based on past observation, its development leaves open the possibility of having missed the Popperian 'Black Swan' (Taleb, 2010). That is, factors responsible for effective treatment delivery (in the addiction field) may have been omitted from the scale as they have not yet been evidenced in the literature. This limitation was unavoidable; all measures are time and context bound. As the field develops, improvements and refinements to the BATS are possible.

The second limitation is in viewing knowledge as a direct or unmediated representation of reality – the “*ontic fallacy*” (Pilgrim, 2014, p.3). The goal of positivist research is to produce objective knowledge that (most likely) reflects reality. While the epistemic fallacy is evident, the error here is ontological. Claiming knowledge of physical phenomena can create false ontological perceptions of reality. With this in mind, the project aimed to understand the contributing factors of therapeutic change in psychotherapy for addiction, without claiming perfect knowledge. It is important to use what appears to work in order to achieve a positive outcome; that is, a tool for evaluating treatment delivery. Given this position, the BATS was developed using multiple methods to continually test and improve our imperfect view of reality (Yardley and Marks, 2004).

## **2.3 Research methods**

The development of the BATS comprised four studies. An overview of each of the studies is provided below.

### **2.3.1 Study 1: Identifying existing measures**

A literature review was conducted to identify fidelity measures that evaluate therapists' delivery of psychological therapies for alcohol and drug use problems. Reviewing what others have done in the past was considered advantageous in developing the BATS (Streiner et al., 2015). The addiction field in particular has been at the forefront in "*addressing the issue of adherence and competence regarding how therapies are being implemented*" (Madson and Campbell, 2006, p.67). The review was conducted in two stages and used a systematic approach. The first stage identified articles from the literature describing the development and/or validation of fidelity measures relevant to the addiction field. The second stage identified relevant fidelity measures from the articles found in stage 1. Consideration was also given to the methods used by the original authors to develop and validate the identified measures; this was important for informing the psychometric work of the BATS.

### **2.3.2 Study 2: Generating an item pool**

Measures identified from the literature review were used as a basis for generating items for the BATS. Generation of items for the BATS was completed in two stages. First, items from the identified measures were analysed using a form of thematic analysis (Braun and Clarke, 2006); items were grouped based on the aspect of therapeutic practice they targeted. Second, the thematic structure (developed in stage 1) was refined, in collaboration with the supervision team, to identify the themes most relevant to the BATS. For example, themes concerning therapy specific techniques were removed as the BATS was designed to be transtheoretical. For the remaining themes, exemplar items were chosen to reflect that aspect of therapeutic practice. These exemplar items were reviewed in Study 3 for potential inclusion in the BATS. At this stage, consideration was also given to scoring; two possible response formats for the BATS were generated using the identified measures as a basis.

### **2.3.3 Study 3: Agreeing the content**

A Delphi study was used to reach a consensus among a group of experts in the fields of addiction and psychotherapy on the content of the BATS (i.e., the items and response

format). The study involved three questionnaires, or rounds, in which participant responses from one round were used to inform the next. The iterative process combined experts' knowledge and opinions to develop a consensus of opinion (McKenna, 1994). As the study progressed, the item pool (generated in Study 2) was reduced and refined. By the end of the third round, a group consensus was obtained. There was agreement that the selected items were comprehensible and important for inclusion in the BATS; items were considered transtheoretical, applicable to therapies widely used for addressing alcohol and drug use problems. At this stage, the first version of the BATS was developed.

#### **2.3.4 Study 4: Testing reliability and validity**

The fourth study developed the BATS further by investigating its psychometric properties. Specifically, the BATS was subjected to convergent validity, and inter-rater reliability analyses. Convergent validity analyses examined the BATS relationship to the Working Alliance Inventory (WAI) and three existing fidelity measures. Inter-rater reliability analyses tested the consistency of measurement on the BATS between two different raters (HC & GT). Convergent validity and inter-rater reliability were examined using secondary analysis of trial data, routine practice data, and process rating data derived from the BATS. Three randomised controlled trials (RCTs) were used in the secondary analysis of trial data: ADAPTA (Watson et al., 2015), AESOPS (Watson et al., 2013a), and UKATT (UKATT Research Team, 2005)<sup>3</sup>. Trial data comprised: recordings of therapy sessions (delivered as part of the trials), process rating data (collected using trial-specific fidelity measures), and WAI outcome data. Routine practice data comprised new recordings of therapy sessions delivered by therapists working at two treatment services: a National Health Service (NHS) specialist addictions service (SAS), and a non-NHS drug and alcohol service (DAS). Process rating data derived from the BATS comprised ratings (collected using the BATS) of therapy sessions selected from the routine practice and trial data.

## **2.4 Ethical considerations**

This section describes the approvals obtained for the project. The main ethical issues are also highlighted, including: consent, confidentiality, data storage, and anonymity of responses.

---

<sup>3</sup> ADAPTA = Addressing Drinking Among Patients: comparing Two Approaches; AESOPS = Alcohol: Evaluating Stepped care in Older Populations Study; UKATT = United Kingdom Alcohol Treatment Trial.

### **2.4.1 Approvals obtained**

Ethical approval for the project was granted by the Yorkshire and the Humber (Leeds West) National Research Ethics Service (NRES) Committee on 17 March 2015 (Reference: 15/YJ/0037). Details of the NRES Committee and Research and Development Department approvals<sup>4</sup> are provided in Appendix A. The project was considered eligible for Clinical Research Network (CRN) support, and was adopted on the National Institute of Health Research CRN Portfolio on 27 January 2015.

### **2.4.2 Protocol changes**

Three amendments were made to the protocol during the research. These changes, relating to Study 3 and Study 4, are summarised below.

#### ***2.4.2.1 Study 3: Agreeing the scale content***

In the original project protocol, items were to be generated from the literature only. A decision was made to include the Delphi exercise because it supported the development of a more robust measure. The Delphi brings together a wide range of knowledge and experience (Murphy et al., 1998), which lends itself to situations where information is lacking or incomplete (Hasson et al., 2000); in this case the specific components of complex therapies are unknown (Michie et al., 2011). The Delphi also enabled content validity to be assessed (Streiner et al., 2015) – the degree to which items on the BATS were relevant and representative of the main techniques for addressing alcohol and drug use problems. Ethical approval for the Delphi study was granted by the School of Medicine Research Ethics Committee at the University of Leeds on 31 October 2016 (Reference: MREC16-008) (Appendix A).

#### ***2.4.2.2 Study 4: Testing reliability and validity***

The protocol originally planned to collect 50 recordings of therapy sessions from the NHS specialist addictions service (SAS) over a 5-month period. Data collection was much slower than expected because the SAS underwent a re-commissioning and re-structuring process; therapists were unable to prioritise the research as a result. In addition, there were a limited number of digital video recorders available, reducing the number of therapists able to partake in the research. Those therapists who did have access to a recorder were

---

<sup>4</sup> The approval documents refer to the BATS as the Addiction Therapist Scale (ATRS). The name of the scale was changed to the BATS during the project.

unfamiliar with their use. This impacted on the quality and number of recordings; for example, inaudible recordings could not be used in the research. To improve data collection, two changes were made to the protocol.

First, the recruitment process was changed to be more inclusive. As it stood, the study could not include recordings of therapy sessions involving friends and family members (FFMs). Therapists working at the SAS reported that clients sometimes attended their appointments with a FFM, particularly for social behaviour and network therapy sessions. In the event of a FFM attending, consent to record the session was needed from both the client and the FFM. The protocol was, therefore, amended to be inclusive of FFMs. The local NRES Committee approved the amendment on 22 May 2015 (Appendix A).

Second, a recruitment site was agreed at a non-NHS alcohol and drug service (DAS), and the recruitment period extended for an additional 17 months. Based on experiences at the SAS, the original aim to collect 50 recordings was revised to a more conservative estimate of at least 20 recordings. The reduced target was not considered problematic, i.e., it did not affect the overall sample size needed for data analysis; recordings of therapy sessions from the three RCTs supplemented the routine practice data. The local NRES Committee approved the protocol changes on 14 September 2015 (Appendix A).

### **2.4.3 Summary of the main issues**

#### **2.4.3.1 Consent**

Participants were recruited in Studies 3 and 4 (**Table 1**). Informed consent was sought before participants took part in the project. Participants were given opportunity to consider a relevant participant information sheet (PIS), ask questions, and decide whether or not they wished to take part. The PIS provided an overview of the relevant study, including an explanation of what taking part would involve, and contact details for the researcher and at least one project supervisor. Two consent forms were returned without a participant signature, any associated data were not used in the project.

The issue of consent was considered in deciding to use the recordings of therapy sessions from the RCTs. Participants involved in the UKATT trial gave consent to use the recordings and associated data for research purposes. Participants involved in the AESOPS and ADAPTA trials gave consent to use the recordings and associated data for quality assurance purposes. The current project falls within this remit as the project aimed to develop a tool

for evaluating treatment delivery in routine practice contexts (i.e., a quality assurance measure).

**Table 1: An overview of the participants recruited**

Study		Participants Recruited
Study 3: Agreeing the content	Delphi exercise	<ul style="list-style-type: none"> <li>• Experts in the fields of addiction and psychotherapy.</li> </ul>
Study 4: Testing reliability and validity	Routine practice data collection	<ul style="list-style-type: none"> <li>• Clients receiving treatment from the SAS or DAS.</li> <li>• Therapists working at the SAS or DAS.</li> <li>• Clients' FFMs.</li> </ul>

DAS = Non-NHS drug and alcohol service; FFMs = Friends and family members; SAS = NHS specialist addictions service.

#### **2.4.3.2 Confidentiality**

Data were treated in confidence. Regarding the recordings from the trial and routine practice data, clients were heard on audio but were not seen in the therapy session recordings. Participants were not identifiable from the information collected in the questionnaires; for example, client data were limited to age, gender, primary problem substance, and treatment history.

#### **2.4.3.3 Data storage**

Participant consent forms were stored in paper format, in a locked cabinet within the secure building of the Leeds Institute of Health Sciences (LIHS). Recordings from the trial data were stored securely at the SAS. These recordings were accessed at the SAS for analysis purposes. Recordings from the routine practice data were transferred onto digital video disc (DVD), which followed routine procedure for recording treatment sessions at the SAS. DVD recordings were temporarily stored securely at the SAS before being moved to LIHS premises. Paper copies of the anonymised questionnaires were stored securely at the LIHS. All electronic data were kept in files on password protected computers.

#### **2.4.3.4 Anonymity of responses**

Completed consent forms were stored separately from other data. Data were identified by a participant identification number. This maintained participant anonymity and enabled the identification of data should participants have requested that their data be withdrawn from the study.

## **2.5 Conclusion**

This project adopted a positivist approach to develop the BATS. There are two main limitations of using a positivist position: the epistemic fallacy, and the ontic fallacy. The first limitation was an inevitable part of the project; all measures are time and context bound. As the field develops, improvements and refinements to the BATS are possible. The second limitation impacted, to an extent, on the methods used to develop the BATS. A multimethod design was used to improve our understanding of how therapy works. Four studies were conducted to develop the BATS. Study 1 identified existing fidelity measures from the literature that evaluate therapists' delivery of psychological therapies for alcohol and drug use problems. Study 2 generated items and response formats for potential inclusion in the BATS using the identified measures as a basis. Study 3 generated a consensus among a group of experts on the content of the BATS, including the items and response format. Study 4 tested the reliability and validity of the BATS. The approvals obtained for the project were described, and the main ethical issues highlighted. The next four chapters describe in more detail each of the four studies undertaken to develop the BATS.



## Chapter 3

### Identifying existing measures

#### 3.1 Introduction

The previous chapter provided an overview of the studies undertaken to develop the BATS. This chapter describes in more detail the first of these four studies. Study 1 was a literature review that aimed to identify fidelity measures that evaluate the delivery of psychological therapies used for alcohol and drug use problems. This chapter describes how the literature was searched, including the inclusion criteria and the methods of searching and synthesising the data. The measures identified from the literature formed the basis of generating items for the BATS; characteristics of the measures are summarised. In addition, this chapter reviews the methods used for examining reliability and validity of the measures; this was important for informing the psychometric work of the BATS in Study 4.

#### 3.2 Method

A literature review was conducted to identify existing fidelity measures that evaluate the delivery of psychological therapies used for addressing alcohol and drug use problems. The review was carried out using a systematic approach, enabling a comprehensive search of the literature. The search was conducted in two stages. First, articles describing the development and/or validation of fidelity measures relevant to the addiction field were identified from the literature. Second, relevant measures (i.e., those that evaluate the delivery of psychological therapies used for alcohol and drug use problems) were identified from the articles found in stage 1. Consideration was also given to the methods used by the original authors to validate the identified measures. This was important for informing the psychometric work of the BATS.

##### 3.2.1 Selection criteria

Criteria for selecting the articles and measures are summarised in **Table 2**.

**Table 2: Criteria for article and measure selection**

Inclusion criteria	Exclusion criteria
<b>Type of articles</b>	
1. Described the development and/or validation of a treatment fidelity measure.	Did not describe the development or validation of a treatment fidelity measure.
2. Published in a peer-reviewed journal.	Not published in a peer-reviewed journal, e.g., conference, dissertation, and thesis abstracts.
<b>Type of measures</b>	
3. Written in the English Language.	Not written in the English Language.
4. Assessed treatment adherence and/or therapist competence.	Did not assess treatment adherence or therapist competence.
5. Assessed therapist behaviours.	Assessed only patient behaviours.
6. Evaluated the delivery of psychological therapies.	Did not evaluate the delivery of psychological therapies, including program fidelity <sup>1</sup> .
7. Evaluated therapies widely used for alcohol and drug use problems, including those originally developed for use in other clinical areas.	Evaluated therapies used for severe mental health or specific physical health problems – treatment too removed from the addiction field.
8. Evaluated therapies not typically used in the addiction field but adapted for alcohol or drug use problems.	Evaluated therapies not typically used in the addiction field and not adapted for alcohol or drug use problems.
9. Evaluated individual therapies.	Evaluated couples, group, or family therapies.
10. Evaluated therapies delivered face-to-face.	Evaluated therapies delivered online or by telephone.
11. Evaluated therapies delivered in the home or in healthcare settings.	Evaluated therapies that are not delivered in the home or in healthcare settings, e.g., schools.
12. Target adult populations (16 years or over).	Target populations under the age of 16 years.

<sup>1</sup> Program fidelity covers multiple interventions and procedures (Bond et al., 2000) of which the delivery of a psychological therapy may be one aspect.

### **3.2.1.1 Type of articles**

Articles describing the development and/or validation of a treatment fidelity measure were eligible for inclusion. Treatment fidelity was defined as the degree to which therapies were delivered as intended (Perepletchikova et al., 2007). In the psychotherapy literature, the extent to which fidelity measures have been validated varies. Development and/or validation articles were, therefore, included in the review. Eligible articles were published

and peer-reviewed. Published articles are indexed in respective databases, enabling a comprehensive search for all relevant literature (Schmucker et al., 2013). Peer-reviewed articles provided some assurance of the quality of the research. The criteria were relatively broad to capture measures accessible in the public domain, measures that are more likely to be used in research and routine practice.

### **3.2.1.2 Type of measures**

Measures eligible for inclusion were written in the English language, the primary language of the United Kingdom (UK).

#### *3.2.1.2.1 Assessment of fidelity*

Measures were included if they evaluated treatment adherence and/or therapist competence, the main components of treatment fidelity. Treatment adherence was defined as the degree to which therapies were delivered as described in the treatment manuals. Therapist competence denoted the level of skill shown by therapists in delivering the therapies. The BATS was designed to inform feedback on clinical skills; therefore, measures focusing on therapist behaviours, rather than the response of the client, were included.

#### *3.2.1.2.2 Target therapies*

Measures targeting psychological therapies were considered for inclusion. A psychological therapy was defined as:

*“...interpersonal treatment that is a) based on psychological principles; b) involves a trained therapist and a client who is seeking help for a mental disorder, problem, or complaint; c) is intended by the therapist to be remedial for the client disorder, problem, or complaint; and d) is adapted or individualized for the particular client and his or her disorder, problem, or complaint.”  
(Wampold and Imel, 2015, p.37)*

The search was not restricted to the addiction field. Measures developed in other clinical areas (e.g., depression) were included if they evaluated therapies also used for substance use problems. Specialised measures that focused on therapies adapted for severe mental health (e.g., schizophrenia) and physical health conditions (e.g., renal disease) were excluded because the treatment for these problems was too removed from the addiction field. However, to capture the breadth of measures, those measures targeting therapies

that are not typically used in the addiction field but were adapted for alcohol or drug use problems were included.

Consideration was also given to the therapy format. Measures were selected if the therapies they evaluated were delivered as individual, face-to-face sessions, involving a therapist and a client (occasionally a friend, family member, or healthcare professional), within the home or healthcare settings. This therapy format is one that is typically offered by addiction services in the UK (Department of Health, 2017). Lastly, measures were considered eligible if the target therapies were suitable for clients aged 16 years and over. This population are able to access adult health services and would normally receive an offer of treatment for substance use problems in the UK (UKATT Research Team, 2005).

### **3.2.2 Search methods**

#### ***3.2.2.1 Stage 1: Identifying articles***

Articles were identified by searching the following databases from the earliest available date to January 2015: PsycINFO (1806 to 2015), Medline (1946 to 2015), and Embase (Embase Classic+Embase, 1947 to 2015). The databases cover different health disciplines but use the same search interface, Ovid. The search strategy was adapted for each database using text words and indexing terms (Appendix B). Other potentially relevant articles were identified by checking the reference lists of articles retrieved from the database searches.

Citations from the database and reference list searches were compiled in an endnote library (EndNote X7). Article titles and abstracts were screened by the researcher. Full manuscripts of potentially relevant articles were retrieved and assessed for inclusion. Articles were referred to the supervision team (GL, BB, and GT) when there was uncertainty about their eligibility for inclusion; for example, when the descriptions of therapies (evaluated by potentially relevant measures) were unclear.

#### ***3.2.2.2 Stage 2: Identifying measures***

Measures that evaluated therapist delivery of psychological therapies for addressing alcohol and drug use problems were identified from the articles found in stage 1. Copies of potentially relevant measures were sourced and assessed for inclusion. Eligibility for inclusion was assessed by the researcher and the supervision team using the study criteria. Where more than one version of a measure was identified, the most recent version was

included subject to meeting the inclusion criteria. The most recent version was not used when the revision precluded the measure's inclusion; for example, measures revised for use with non-English speakers. In such instances, a previous version of the measure was included. Measures were sourced: i) directly from the development and/or validation articles, ii) by searching the internet, or iii) by contacting the authors of relevant articles – the authors contacted about particular measures were also asked if they had developed other measures that met the inclusion criteria. Measures that could not be obtained were excluded.

### **3.2.3 Summarising the literature**

#### ***3.2.3.1 Characteristics of the identified measures***

Characteristics of the measures that met the inclusion criteria were summarised in a table. The information presented included: the clinical area, purpose (e.g., for research or clinical supervision), target therapies (e.g., motivational interviewing), assessment of fidelity (treatment adherence and/or therapist competence), number of items, and item scoring (how the majority of items were scored).

#### ***3.2.3.2 Methods of validating the identified measures***

Psychometric properties of the identified measures were summarised in a table. The information presented focused on the development of the measures, and the main validation analyses. The evidence in support of the identified measures was reviewed. This included a description of the methods used to validate the measures, and a brief evaluation. The methods described were taken from articles found in stage 1 associated with the identified measures.

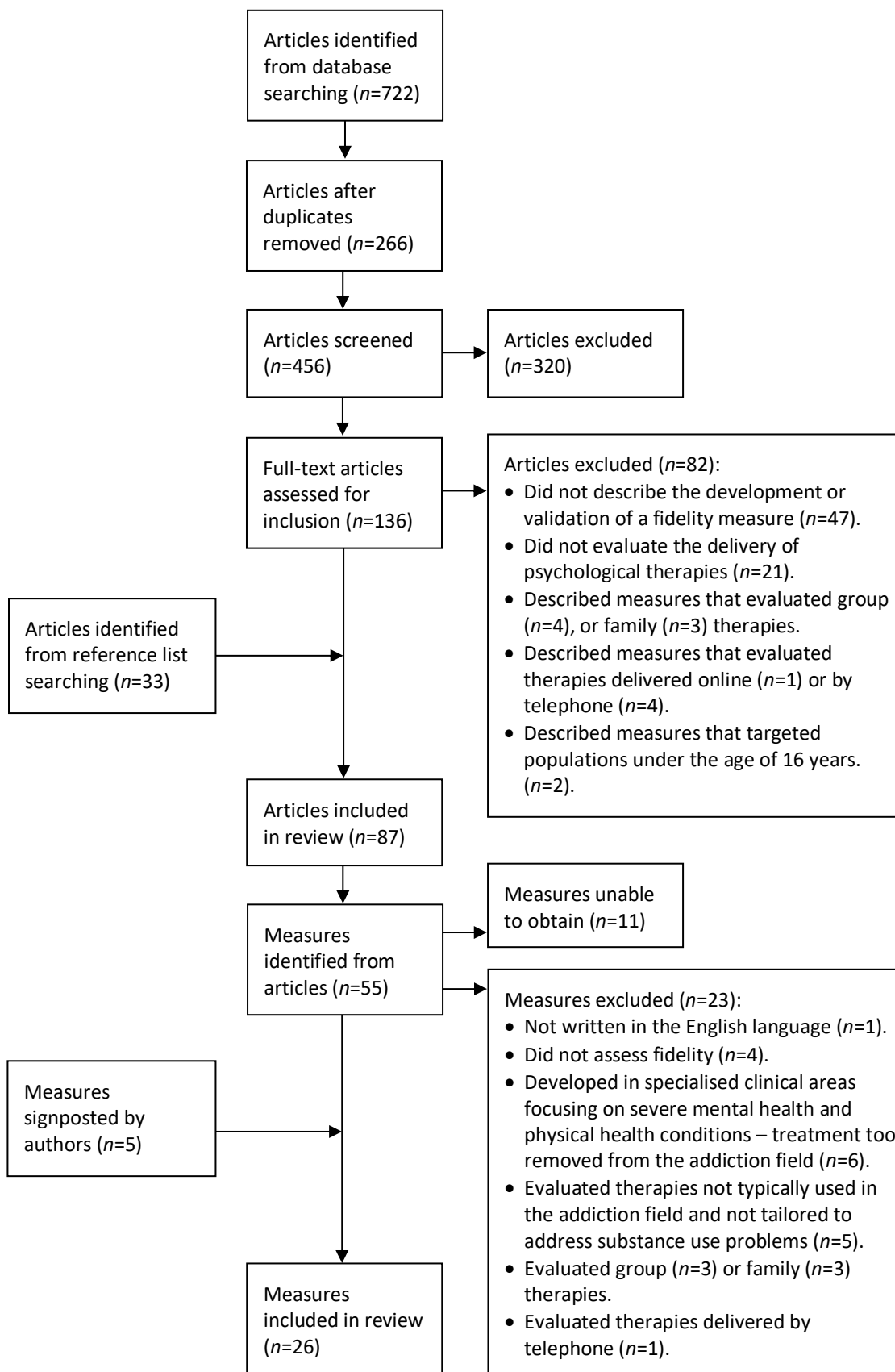
## **3.3 Results**

### **3.3.1 Outcome of searches**

The selection of articles and measures are summarised in **Figure 1**.

#### ***3.3.1.1 Stage 1: Identifying articles***

Searching the healthcare databases identified 722 records, of which 266 were found to be duplicates (identified in more than one database): Embase identified 299 records, Medline 153 records, and PsycINFO 270 records. Following the removal of duplicates, article screening excluded a further 320 records. Full manuscripts of 136 records were retrieved and assessed for inclusion. Eighty two records were excluded. There were two main



**Figure 1: Flowchart illustrating the selection of measures included in the review**

reasons for exclusion: i) articles did not describe the development and/or validation of a fidelity measure (e.g., review papers and commentaries;  $n=47$ ), and ii) articles were concerned with measures that did not evaluate psychological therapies (e.g., community-based interventions for pain;  $n=21$ ). Reference list searching identified an additional 33 relevant articles, resulting in 87 articles included in the review.

### **3.3.1.2 Stage 2: Identifying measures**

Fifty-five measures were identified from the 87 articles found in stage 1. Copies of the measures were assessed for inclusion, of which 23 were excluded. The main reasons for exclusion were: i) measures developed in specialised areas focusing on severe mental health and physical health conditions, the treatment for these conditions was considered too removed from the addiction field (e.g., psychosis;  $n=6$ ), and ii) measures evaluated psychological therapies not typically used in the addiction field and were not adapted for alcohol or drug use problems (e.g., brief psychodynamic investigation;  $n=5$ ). Copies of 11 measures could not be obtained and were excluded. Five additional measures were obtained from contacting the authors<sup>5</sup>. Thus, a total of 26 measures were included in the review. Excluded measures are summarised in Appendix B.

### **3.3.2 Characteristics of the identified measures**

The 26 identified measures are summarised in **Table 3**, providing an overview of the measures' clinical focus, purpose, target therapies, number of items, and item scoring.

#### **3.3.2.1 Clinical focus**

Substance use was the main clinical focus for 15 measures (e.g., UKATT PRS; Middleton et al., 2001)<sup>6</sup>; nine of these specifically targeted drug abuse and dependence ( $n=5$ ) (e.g., CE Therapist Rating Form; Carroll et al., 1999)<sup>7</sup>, problem drinking ( $n=3$ ) (e.g., ADAPTA PRS;

---

<sup>5</sup> Authors contacted about measures identified from the literature were also asked if they had developed other measures that met the inclusion criteria. Five such measures were signposted by the authors: i) ADAPTA PRS = ADAPTA (Addressing Drinking Among Patients: comparing Two Approaches) Process Rating Scale (Tober and Crosby, 2014), ii) AESOPS PRS = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale (Tober and Crosby, 2011), iii) CE Therapist Rating Form = Compliance Enhancement Therapist Adherence/ Competence Rating Form (Carroll et al., 1999), iv) CM Clinician Rating Form = Contingency Management Clinician Adherence/Competence Rating Form (Petry and Stitzer, 2002), and v) GROMIT = Global Rating of Motivational Interviewing Therapist (Moyers, 2004).

<sup>6</sup> UKATT PRS = UKATT (United Kingdom Alcohol Treatment Trial) Process Rating Scale

<sup>7</sup> CE Therapist Rating Form = Compliance Enhancement Therapist Adherence/ Competence Rating Form

Tober and Crosby, 2014)<sup>8</sup>, and prenatal substance use (IAC Treatment Fidelity Instrument; Torrey, 2011)<sup>9</sup>. Mental health was the main focus for six measures, including: depression ( $n=3$ ) (e.g., SPRS; Shapiro and Startup, 1990)<sup>10</sup>, parasuicide (MACT Rating Scale; Davidson et al., 2004)<sup>11</sup>, post-traumatic stress disorder (CBT for PTSD Fidelity Scale; Lu et al., 2012)<sup>12</sup>, and dual diagnosis (i.e., comorbidity of substance use and mental health problems, in this case psychosis) (MI-CTS; Haddock et al., 2012)<sup>13</sup>. The remaining five measures focused on the therapeutic approach, rather than particular health-related problems. The approaches included: cognitive therapy (CTS-R; Blackburn et al., 2001b)<sup>14</sup>, motivational interviewing (MISTS Revised; Madson and Loignon, 2007)<sup>15</sup>, psychotherapy (CPPS; Hilsenroth et al., 2005)<sup>16</sup>, and general behaviour change ( $n=2$ ) (e.g., TPRS; Fisher et al., 2000)<sup>17</sup>.

### **3.3.2.2 Purpose**

The majority of measures were developed for research purposes ( $n=24$ ) (e.g., TSF-ACES; Campbell and Guydish, 2012)<sup>18</sup>. The remaining two measures applied to training, and supervision settings (CTS-R, Blackburn et al., 2001b; MISTS Revised, Madson and Loignon, 2007). Of the 24 research measures, six were developed for use in additional contexts, including: training, supervision, and self-reflective practice (e.g., CM Clinician Rating Form; Petry and Stitzer, 2002)<sup>19</sup>.

### **3.3.2.3 Target therapies**

The majority of measures were designed to evaluate therapist adherence and/or competence in delivering one specific treatment modality ( $n=17$ ) (e.g., TSF-ACES; Campbell and Guydish, 2012). A minority of measures targeted two ( $n=3$ ) or three ( $n=2$ ) different

---

<sup>8</sup> ADAPTA PRS = ADAPTA (Addressing Drinking Among Patients: comparing Two Approaches) Process Rating Scale.

<sup>9</sup> IAC Treatment Fidelity Instrument = I Am Concerned (a brief opportunistic intervention for prenatal substance use) Fidelity Instrument.

<sup>10</sup> SPRS = Sheffield Process Rating Scale.

<sup>11</sup> MACT Rating Scale = Manual Assisted Cognitive Therapy Rating Scale.

<sup>12</sup> CBT for PTSD Fidelity Scale = Cognitive-Behavioral Therapy for Post-Traumatic Stress Disorder Fidelity Scale.

<sup>13</sup> MI-CTS = Integrated Motivational Interviewing and Cognitive Behavioural Therapy Fidelity Scale.

<sup>14</sup> CTS-R = Cognitive Therapy Scale – Revised.

<sup>15</sup> MISTS Revised = Motivational Interviewing Supervision and Training Scale Revised.

<sup>16</sup> CPPS = Comparative Psychotherapy Process Scale.

<sup>17</sup> TPRS = Therapy Process Rating Scale.

<sup>18</sup> TSF-ACES = Twelve Step Facilitation Adherence Competence Empathy Scale.

<sup>19</sup> CM Clinician Rating Form = Contingency Management Clinician Adherence/Competence Rating Form.



**Table 3: Characteristics of the identified fidelity measures**

Measure (Author, Year)	Clinical Area	Purpose	Target Therapies	Assessment of Fidelity	No. Items	Item Scoring
1. ACS-IDCCD (Mercer et al., 1995)	Cocaine dependence	Research	Individual drug counselling (IDC)	Adherence (frequency), and competence (quality)	43	7-point rating scales scoring item frequency, and quality.
2. ACS-SEC (Barber, 1997)	Cocaine dependence	Research	Supportive-expressive therapy (SET)	Adherence (frequency), and competence (appropriateness and quality)	82	7-point rating scales scoring item frequency, appropriateness, and quality.
3. ADAPTA PRS (Tober and Crosby, 2014)	Problem drinking	Research	An alcohol focused intervention (AF), a healthy living intervention (HL)	Adherence (extensiveness), and competence (quality)	15	5-point rating scales scoring item frequency, and quality. Dichotomous ratings for session content item ("yes/no").
4. AESOPS PRS (Tober and Crosby, 2011)	Hazardous drinking in older adults	Research	Brief advice (BA), behaviour change counselling (BCC), motivational enhancement therapy (MET)	Adherence (extensiveness), and competence (quality)	19	5-point rating scales scoring item extensiveness, and quality. Dichotomous ratings for one item covering session content ("yes/no").
5. BAS (Pantalon et al., 2012)	Harmful and hazardous drinking	Research and training	Brief Negotiation Interview (BNI)	Adherence	21	Dichotomous ratings ("yes/no").
6. BECCI (Lane, 2002)	Behaviour change	Research and training	BCC	Competence (extensiveness)	11	5-point rating scales scoring item extensiveness.
7. CBT for PTSD Fidelity Scale (Lu et al., 2012)	Serious mental illness and post-traumatic stress disorder (PTSD)	Research	Cognitive behavioural therapy (CBT)	Combined adherence and competence (quality)	17	5-point rating scales scoring item quality.

Measure (Author, Year)	Clinical Area	Purpose	Target Therapies	Assessment of Fidelity	No. Items	Item Scoring
8. CBT Therapist Checklist <sup>1</sup> (Carroll, 1997)	Cocaine dependence	Research, supervision, and self-reflective practice	CBT	Adherence (extensiveness)	28	7-point rating scales scoring item extensiveness.
9. CE Therapist Rating Form (Carroll et al., 1999)	Drug abuse and dependence	Research, supervision, and self-reflective practice	Compliance enhancement (CE)	Adherence (extensiveness), and competence (quality)	28	7-point rating scales scoring item extensiveness, and quality.
10. CM Clinician Rating Form (Petry and Stitzer, 2002)	Substance use	Research and training	Contingency management (CM)	Adherence (extensiveness), and competence (quality)	13	10 items scored on 7-point rating scales assessing extensiveness, and quality. Three general skilfulness items scored for quality only.
11. CPPS (Hilsenroth et al., 2005)	Psychotherapy techniques	Research, clinical practice, and training	Psychodynamic-interpersonal (PI) and cognitive-behavioural (CB) treatments	Quantity (term adherence not used, CPPS rates features of PI and CB treatments)	20	7-point rating scales scoring the extent to which item behaviours were characteristic of the session.
12. CSPRS-6 (SPR Project Staff, 1984)	Depression	Research	CBT, interpersonal therapy (IPT), clinical management (ClinM)	Adherence (extensiveness)	96	7-point rating scales scoring item extensiveness.
13. CTACS <sup>2</sup> (Liese et al., 1995)	Cocaine dependence	Research	Cognitive Therapy (CT)	Adherence, and competence (appropriateness and quality)	28	7-point rating scales scoring item adherence, appropriateness, and quality. Quality ratings include item specific anchors.

Measure (Author, Year)	Clinical Area	Purpose	Target Therapies	Assessment of Fidelity	No. Items	Item Scoring
14. CTS-R (Blackburn et al., 2001b)	Clients receiving cognitive therapy (CT)	Training	CT	Competence (therapist skill, appropriateness, and suitability of methods)	12	7-point rating scales scoring competence with item specific anchors.
15. GROMIT (Moyers, 2004)	Substance use	Research	Motivational interviewing (MI)	Adherence (agreement)	15	7-point rating scales scoring occurrence of therapist behaviours.
16. IAC Treatment Fidelity Instrument (Torrey, 2011)	Prenatal substance use	Research	I Am Concerned (IAC) brief opportunistic intervention	Adherence, and competence	18	3-point and 5-point rating scales scoring adherence and competence. Both scales rate level of agreement.
17. ITRS (Martino et al., 2009b)	Substance use	Research	MI	Adherence (frequency and extensiveness), and competence (skilfulness)	42	25 items scored on 7-point rating scales assessing adherence, and competence. 17 items (general therapist and client behaviours) scored on one 7-point rating scale; item specific descriptive anchors.
18. MACT Rating Scale (Davidson et al., 2004)	Parasuicide	Research	Manual assisted cognitive therapy (MACT)	Competence (skilfulness, interpersonal effectiveness, and general adherence to therapy model)	11	7-point rating scales scoring competence with item specific anchors.
19. MI-CTS (Haddock et al., 2012)	Psychosis and substance use	Research	Integrated MI and CBT (MI-CBT)	Adherence (extensiveness)	19	3-point rating scales scoring adherence. One item asks for a list of the MI and CBT techniques delivered.

Measure (Author, Year)	Clinical Area	Purpose	Target Therapies	Assessment of Fidelity	No. Items	Item Scoring
20. MISTS Revised <sup>3</sup> (Madson and Loignon, 2007)	Health related problems where MI is incorporated	Training and supervision	MI and its derivatives, e.g., MET	Competence (therapist skill)	20	Global ratings made on a 7-point rating scale using item specific anchors to score therapist skill.
21. MTRS (DeRubeis et al., 1982)	Depression	Research	CBT, and IPT	Adherence	32	9-point rating scales scoring adherence with item specific descriptive end anchors.
22. SPRS (Shapiro and Startup, 1990)	Depression	Research	Exploratory therapy (ET)	Adherence (extensiveness)	59	7-point rating scales scoring extensiveness.
23. TPRS (Fisher et al., 2000)	Behaviour change	Research	Psychotherapy	Adherence (frequency, extensiveness, agreement, percentage), therapist skill (agreement)	59	5-point rating scales scoring frequency, extensiveness, percentage (e.g., client level of verbal activity), and agreement.
24. TSF-ACES <sup>4</sup> (Campbell and Guydish, 2012)	Substance use	Research	Twelve-step facilitation (TSF)	Adherence (extensiveness), and competence (skill)	57	7-point rating scales scoring extensiveness, and therapist skill.
25. UKATT PRS <sup>5</sup> (Middleton et al., 2001)	Substance use	Research	MET, and social behaviour and network therapy (SBNT)	Adherence (frequency), and competence (skilfulness)	28	5-point rating scales scoring extensiveness, and quality. Dichotomous ratings for session content item and the 7 additional items (session structure and client behaviours) (“yes/no”).

Measure (Author, Year)	Clinical Area	Purpose	Target Therapies	Assessment of Fidelity	No. Items	Item Scoring
26. YACSII (Nuro et al., 2005)	Substance use	Research	ClinM, IPT, TSF, CBT, and MI	Adherence (frequency & extensiveness), and competence (skill level)	68	7-point rating scales scoring combined frequency and extensiveness, and skill level.

<sup>1</sup> The treatment manual contains two measures: CBT Therapist Checklist, and CBT Rating Scale (Carroll, 1997). Both include the same items but evaluate different attributes: the validated Checklist assesses adherence only (Carroll et al., 1998b), the Rating Scale evaluates adherence and competence.

<sup>2</sup> CTACS comprised 28 items/questions: 25 items focusing on therapist behaviours, and three additional questions on client difficulty, therapist strengths and weaknesses; all 28 items/questions were counted.

<sup>3</sup> MISTS included two components: i) a global rating scale on key aspects of motivational interviewing, and ii) a behaviour count of therapist responses (Madson et al., 2005). The global rating scale was revised, and named: MISTS Revised (Madson and Loignon, 2007). Therapist responses are counted using a behaviour tracking form; the form was not included in the manual nor referred to in the validation article.

<sup>4</sup> TSF-ACES comprised four session-specific rating forms: one for groups (15 items) and three for individual sessions (57 items). The number of items for the individual session rating forms was counted; BATS was designed to evaluate individual therapies.

<sup>5</sup> UKATT PRS comprised 35 items/questions: 28 items focusing on therapist behaviours, and seven additional questions on the client and the session structure; all 35 items/questions were counted.

No. Items = Number of questions or items.

**ACS-IDCCD** = Adherence/Competence Scale for Individual Drug Counseling for Cocaine Dependence; **ACS-SEC** = Adherence/Competence Scale for Supportive-Expressive Therapy for Cocaine Dependence; **ADAPTA PRS** = ADAPTA (Addressing Drinking Among Patients: comparing Two Approaches) Process Rating Scale; **AESOPS PRS** = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale; **BAS** = Brief Negotiation Interview Adherence Scale; **BECCI** = Behaviour Change Counselling Index; **CBT for PTSD Fidelity Scale** = Cognitive-Behavioral Therapy for Post-Traumatic Stress Disorder Fidelity Scale; **CBT Therapist Checklist** = Cognitive Behavioral Therapy Therapist Checklist; **CE Therapist Rating Form** = Compliance Enhancement Therapist Adherence/Competence Rating Form; **CM Clinician Rating Form** = Contingency Management Clinician Adherence/Competence Rating Form; **CPPS** = Comparative Psychotherapy Process Scale; **CSPRS-6** = Collaborative Study Psychotherapy Rating Scale – Form 6; **CTACS** = Cognitive Therapy Adherence and Competence Scale; **CTS-R** = Cognitive Therapy Scale – Revised; **GROMIT** = Global Rating of Motivational Interviewing Therapist; **IAC Treatment Fidelity Instrument** = I Am Concerned Treatment Fidelity Instrument; **ITRS** = Independent Tape Rater Scale; **MACT Rating Scale** = Manual Assisted Cognitive Therapy Rating Scale; **MI-CTS** = Integrated Motivational Interviewing and Cognitive Behavioural Therapy Fidelity Scale; **MISTS Revised** = Motivational Interviewing Supervision and Training Scale Revised; **MTRS** = Minnesota Therapy Rating Scale; **SPRS** = Sheffield Psychotherapy Rating Scale; **TPRS** = Therapy Process Rating Scale; **TSF-ACES** = Twelve Step Facilitation Adherence Competence Empathy Scale; **UKATT PRS** = UK Alcohol Treatment Trial Process Rating Scale; **YACSII** = Yale Adherence and Competence Scale Second Edition.

therapies (e.g., UKATT PRS, Middleton et al., 2001; CSPRS-6, SPR Project Staff, 1984)<sup>20</sup>. The remaining four measures encompassed a much broader range of therapies. For example, the CPPS was designed to rate key features of psychodynamic-interpersonal and cognitive-behavioural therapies (Hilsenroth et al., 2005).

There were twenty five different therapies covered by the measures, a brief description of each therapy is provided in Appendix B. The most frequently evaluated therapies were: cognitive-behavioural therapy (CBT;  $n=5$ ) (e.g., CBT Therapist Checklist; Carroll, 1997)<sup>21</sup>, motivational interviewing (MI;  $n=3$ ) (e.g., MISTS Revised; Madson and Loignon, 2007), and clinical management, also called compliance enhancement ( $n=3$ ) (e.g., YACSII; Nuro et al., 2005). **Figure 2** shows that some therapies were based on those previously developed (and are, therefore, related to them). For example, there were three derivatives of MI, including: behaviour change counselling (BCC;  $n=2$ ) (e.g., BECCI; Lane, 2002)<sup>22</sup>, motivational enhancement therapy (MET;  $n=2$ ) (e.g., AESOPS PRS; Tober and Crosby, 2011)<sup>23</sup>, and the brief negotiation interview (BNI;  $n=1$ ) (BAS; Pantaloni et al., 2012)<sup>24</sup>. Unlike other therapies, integrated MI-CBT was developed from two treatment modalities, MI and CBT (MI-CTS; Haddock et al., 2012).

#### **3.3.2.4 Assessment of fidelity**

Most measures assessed one component of treatment fidelity ( $n=13$ ); nine measures evaluated treatment adherence (e.g., GROMIT; Moyers, 2004)<sup>25</sup>, and four measures assessed therapist competence (e.g., CBT for PTSD Fidelity Scale; Lu et al., 2012). Therapist adherence refers to the extent to which therapies were delivered as described in the treatment manuals, whereas therapist competence denotes the level of skill shown by therapists in delivering the therapies. Twelve measures assessed both adherence and competence (e.g., AESOPS PRS; Tober and Crosby, 2011), two of these measures also

---

<sup>20</sup> CSPRS-6 = Collaborative Study Psychotherapy Rating Scale – Form 6.

<sup>21</sup> CBT Therapist Checklist = Cognitive Behavioral Therapy Therapist Checklist.

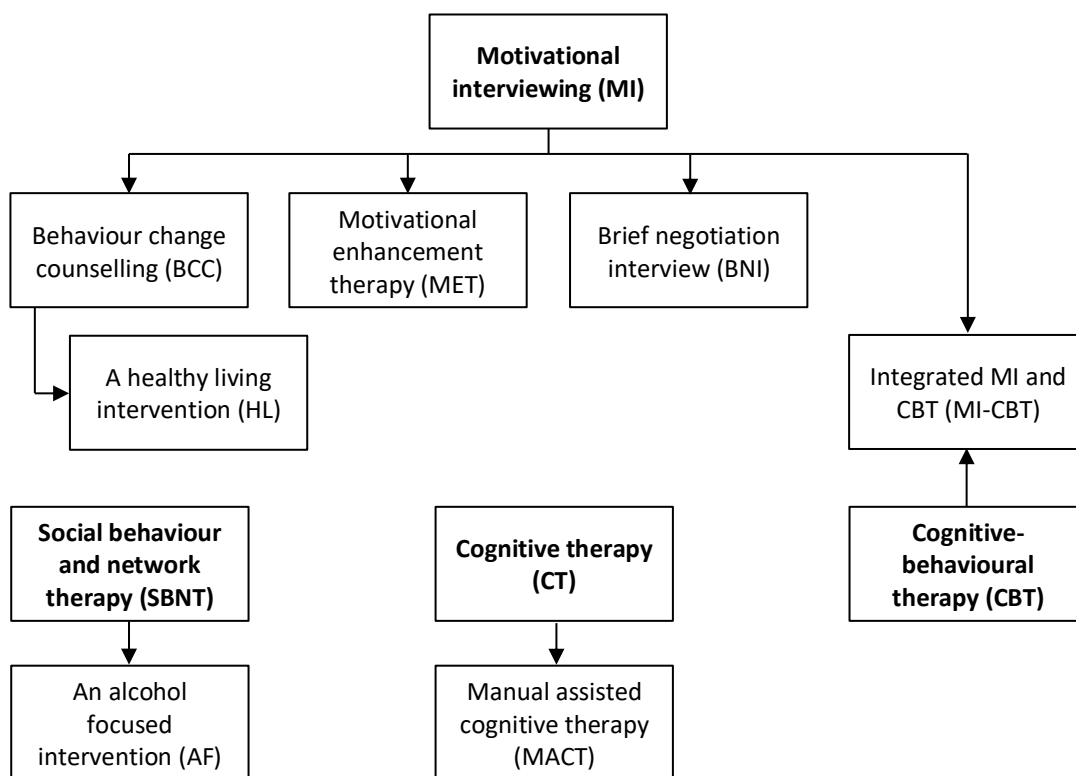
<sup>22</sup> BECCI = Behaviour Change Counselling Index.

<sup>23</sup> AESOPS PRS = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale.

<sup>24</sup> BAS = Brief Negotiation Interview Adherence Scale.

<sup>25</sup> GROMIT = Global Rating of Motivational Interviewing Therapist.

evaluated the appropriateness of item behaviours (e.g., CTACS; Liese et al., 1995)<sup>26</sup>. The components of fidelity assessed by the MTRS<sup>27</sup> were not specified (DeRubeis et al., 1982).



**Figure 2: Relationships between therapies covered by the identified measures**

Treatment adherence was assessed in five ways. First, by rating item extensiveness, the level of depth or detail with which therapists delivered item behaviours ( $n=11$ ) (e.g., CBT Therapist Checklist; Carroll, 1997). Second, by rating item frequency, the number of instances the therapists carried out item behaviours ( $n=5$ ) (e.g., ACS-IDCCD; Mercer et al., 1995)<sup>28</sup>. Third, by rating item frequency and extensiveness, the amount of time and attention given by therapists in delivering item behaviours ( $n=2$ ) (e.g., YACSII; Nuro et al., 2005)<sup>29</sup>. Fourth, by rating item agreement, the level of agreement concerning the occurrence of item behaviours ( $n=2$ ) (e.g., IAC Treatment Fidelity Instrument, which differentiated between adherence behaviours and more complex competence behaviours; Torrey, 2011). Lastly, by rating items dichotomously, whether or not item behaviours were carried out ( $n=1$ ) (BAS; Pantalon et al., 2012).

<sup>26</sup> CTACS = Cognitive Therapy Adherence and Competence Scale.

<sup>27</sup> MTRS = Minnesota Therapy Rating Scale.

<sup>28</sup> ACS-IDCCD = Adherence/Competence Scale for Individual Drug Counseling for Cocaine Dependence.

<sup>29</sup> YACSII = Yale Adherence and Competence Scale Second Edition.

Competence was assessed in four ways. First, by rating the quality or skilfulness with which therapists performed item behaviours ( $n=13$ ) (e.g., ITRS; Martino et al., 2009b). Second, by rating the level of skill and appropriateness of the item behaviours delivered by the therapists ( $n=1$ ) (CTS-R; Blackburn et al., 2001b). Third, by rating three aspects of therapist competence, including skilfulness, interpersonal effectiveness, and general adherence ( $n=1$ ) (MACT Rating Scale; Davidson et al., 2004). Lastly, by rating the level of agreement concerning the occurrence of complex behaviours ( $n=1$ ) (IAC Treatment Fidelity Instrument; Torrey, 2011).

### **3.3.2.5 Number of items**

There was a mean of 33 items across the identified measures. The shortest two measures comprised 11 items (BECCI, Lane, 2002; MACT Rating Scale, Davidson et al., 2004), and the longest 96 items (CSPRS-6; SPR Project Staff, 1984). Measures developed for sole use in research (e.g., ACS-SEC; Barber, 1997)<sup>30</sup> were generally longer than those developed for other contexts (including research) (e.g., CE Therapist Rating Form; Carroll et al., 1999); there was a mean of 39 items for the research only measures ( $n=18$ ; range: 11 to 96), and a mean of 19 items for the other measures ( $n=8$ ; range: 11 to 28).

### **3.3.2.6 Item scoring**

The next two sections (response formats, and response options) summarise the main scoring methods used by the identified measures.

#### **3.3.2.6.1 Response formats**

The measures used four main response formats: adjectival scales, Likert scales, ordered-categorical scales, and dichotomous scoring. Adjectival scales comprise unipolar response options, ranging from none or a little of the attribute to the maximum amount (Streiner et al., 2015). These scales are measured at the ordinal level; that is, the response options have an order but the interval between options is not necessarily equal (Jamieson, 2004). Seventeen measures used adjectival scales to assess treatment adherence ( $n=17$ ), and the appropriateness of item behaviours ( $n=2$ ). In relation to adherence, Likert scales were uncommon; only three of the 21 measures that assessed treatment adherence used this response format. Likert scales are similar to adjectival scales but they include bipolar

---

<sup>30</sup> ACS-SEC = Adherence/Competence Scale for Supportive-Expressive Therapy for Cocaine Dependence.



response options (Streiner et al., 2015) (Table 4). Likert scales were frequently used to rate therapist competence; eleven of the 16 measures assessing competence used Likert scales.

**Table 4: Response formats used in the AESOPS PRS**

Assessment of Fidelity (Response Format)	Response Options				
Adherence (Adjectival scale)	0 Not at all	1 A little	2 Somewhat	3 Considerably	4 Extensively
Competence (Likert scale)	0 Very poor	1 Poor	2 Good enough	3 Well	4 Very well

**AESOPS PRS** = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale.

Ordered-categorical scales were used by the remaining five measures that assessed therapist competence. Ordered-categorical scales are Likert-type scales that contain item-specific descriptive anchors:

*“Quality rating (“NA” if appropriately not done):*

- 0 *The therapist seemed unaware of the patient's agenda.*
- 2 *The therapist elicited agenda items that were vague or incomplete.*
- 4 *The therapist elicited agenda items and attempted, with some success, to prioritize these and follow agenda.*
- 6 *The therapist set an excellent, comprehensive agenda, identified important target problems, prioritized, and followed agenda.” (CTACS; Liese et al., 1995)*

ACS-SEC was unique in that the measure also provided an overarching Likert scale (“very poorly/acceptably/very well”) to guide ratings made on the item-specific ordered-categorical scales (Barber, 1997). Dichotomous scoring was the least common response format; only one measure provided binary options (“Yes/No”) for each item to assess treatment adherence (BAS, Pantaloni et al., 2012; DeVellis, 2017). The MTRS was not included in the above summary; information on the selected response format(s) was limited: *“items were written in Likert-type format”* (DeRubeis et al., 1982, p.246).

### 3.3.2.6.2 Response options

The majority of measures scored items on 5-point ( $n=6$ ) and 7-point ( $n=15$ ) scales. The points were all labelled on the 5-point scales (e.g., TPRS; Fisher et al., 2000), excepting the end-anchored competence (quality) scale on the UKATT PRS (Middleton et al., 2001). Of the 7-point scales, most were labelled every other point (e.g., CSPRS-6; SPR Project Staff, 1984), providing four labelled anchors; all seven points were labelled in only four measures

(e.g., YACSII; Nuro et al., 2005). Other categories were used, including: nine (MTRS; DeRubeis et al., 1982), six (TSF-ACES; Campbell and Guydish, 2012), three (MI-CTS; Haddock et al., 2012), and two response options (BAS; Pantalon et al., 2012). In these four measures, all points were labelled excepting the end-anchored 9-point scales in the MTRS (DeRubeis et al., 1982). The remaining measure used different length rating scales (IAC Treatment Fidelity Instrument; Torrey, 2011); specifically, adherence items were scored 3-point scales, and competence items on 5-point scales; points on both scales were all labelled.

Ten measures included a 'not applicable' response option. In six measures the option was given for most items; specifically, for rating item competence when the specified behaviour was not observed during the session (e.g., ITRS; Martino et al., 2009b). Three measures included the 'not applicable' option for a minority of items relating to treatment adherence. In these cases, the option was available when the specified behaviour was not applicable to the session context (e.g., BECCI; Lane, 2002). The remaining measure included the response option for rating item competence, and for rating one adherence item (IAC Treatment Fidelity Instrument; Torrey, 2011).

### **3.3.3 Methods of validating the identified measures**

So far, this review has focused on the characteristics of the identified measures. Consideration was also given to the methods used by the original authors to develop and validate the measures. This was important for informing the psychometric work of the BATS. Twenty one articles reported the measures' psychometric properties<sup>31</sup> (Appendix B). Thus, the next part of this review provides an overview of the main types of validity and reliability, with a discussion on the articles' chosen analytic methods.

#### **3.3.3.1 Validity**

Validity reflects "*the ability of a scale to measure what it is supposed to measure*" (Bowers et al., 2014, p.109). The articles evidenced five main types of validity: construct validity, content and face validity, convergent validity, discriminant validity, and sensitivity for change.

---

<sup>31</sup> Development and/or validation articles for the MISTS Revised and YACSII were not identified. The articles found related to previous versions of the measures; the psychometric properties of MISTS and YACS may not generalise to different treatments, populations, and scale versions. However, the analytic methods were summarised to help inform the validation of the BATS.

### 3.3.3.1.1 Construct validity

Construct validity refers to a measure's ability to assess the underlying constructs it purports to measure (DeVellis, 2017). Constructs are theoretical concepts that explain the relationships among a set of items (Streiner et al., 2015). Construct validity differs from other forms of validity because it integrates "*any evidence that bears on the interpretation or meaning of the test scores*" (Messick, 1995, p.742). In other words, it is comprehensive, subsuming other forms of validity, including: content, convergent and discriminant validity (Strauss and Smith, 2009). It is unsurprising, therefore, that there are multiple ways to demonstrate construct validity. Four such methods reported in eight of the articles were: principal components analysis (PCA), exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and a construct explication exercise.

PCA, described in three of the articles (e.g., UKATT PRS; Tober et al., 2008), is a method for transforming a relatively large number of variables (i.e. items) into a smaller set of components (Field, 2017). The components describe the linear combinations of the items, and so explain the maximum amount of total variance in the data (DeCoster, 1998; Field, 2017). Total variance includes two parts: common variance, and unique variance. Common variance is the proportion of total variance for an item measurement that is shared with other items included in the analysis (known as communality) (Field, 2017). Unique variance is the amount of variance specific to a particular item, including both systematic and random (error) variability (O'Rourke and Hatcher, 2013). Because PCA analyses total variance, it is argued that the method is inappropriate for demonstrating construct validity (Henson and Roberts, 2006). Indeed, PCA can provide a "*misleading picture of the factor structure underlying the data*" (O'Rourke and Hatcher, 2013, p.52). Thereby, limiting replication in other samples and confirmation through CFA (Wetzal, 2012). The method does have value in the early stages of scale development; for example, the number of items on the MTRS were reduced (based on the results of a PCA) to better reflect the main therapist interventions and strategies of the two different treatments (DeRubeis et al., 1982).

EFA is also a data reduction technique, and follows similar analytic procedures to PCA. However, there are two key differences between the methods. First, EFA assumes that the variables (or items) are based on a number of underlying factors (DeCoster, 1998). The reverse is true in PCA: the components are based on the item measurements. Second, unlike PCA, EFA discriminates between common and unique variance (Costello and

Osborne, 2005; Henson and Roberts, 2006). Because EFA focuses on the shared variance, it is preferred to PCA for exploring possible factors that underlie a set of items (DeVellis, 2017). Appendix B illustrates the differences between the methods (including CFA). Two articles described the use of EFA to evidence construct validity. The first article used an EFA for dichotomous variables (Pantalon et al., 2012). While appropriate for the data collected (items on the BAS were scored dichotomously), it was not clear whether EFA or PCA had been conducted; the article included terms relating to both methods, e.g., 'factors' and 'components'. The second article provided a comprehensive description of the EFA procedure, including information on the rotation method, and criteria for factor retention (Shapiro and Startup, 1992). Information on sample size adequacy was omitted from both articles. Reporting a measure, such as the Kaiser-Meyer-Olkin statistic, would have been advantageous in gauging the reliability of the analyses (Field, 2017). That being said, both articles included samples greater than 300, increasing the likelihood of a stable factor solution (Field, 2017).

EFA provides relatively weak evidence of construct validity because the researchers have no *a priori* hypotheses about the nature of the underlying factor structure (Streiner et al., 2015); all items are related to every factor (Appendix B). In contrast, CFA is a hypothesis-testing technique, used to examine a predefined pattern of relationships among a set of items based on theory or previous research (DeVellis, 2017; DeCoster, 1998); researchers "*can specify which items comprise each factor*" (Streiner et al., 2015, p.379). Thus, CFA can provide stronger support of construct validity than EFA.

Three articles used CFA to evidence construct validity of the measures. Positive reporting practices were found for justifying the model formulation, and evaluating the model fit. For example, all three articles used several goodness-of-fit indices to evidence "*whether the hypothesized model was a good fit to the observed data*" (Schreiber et al., 2006, p.327); cut off criteria for the indices were specified, e.g., chi-square degrees of freedom ratios <2. However, information on data screening and other CFA procedures was lacking. For example, only one article reported the chosen estimation method (ITRS; Martino et al., 2008). Estimation methods fit the model to the data; they are required to "*obtain estimates of the relationships among variables in the mathematics model*" (Schreiber et al., 2006, p.336). Martino et al. (2008) used the maximum likelihood estimation, which assumes a multivariate normal distribution. However, checks for violations of normality were not reported (in any of the three articles). Such basic information is important for

evaluating the appropriateness of the researchers' analytic decisions and the validity of the results (Jackson et al., 2009).

Lane et al. (2005) reported a different approach to examining construct validity of the BECCI. A construct explication exercise was conducted, which described the relationship between item behaviours and the target therapy, BCC. This exercise relied on the subjective impressions of the researchers (DeVellis, 2017); the reliability of their judgements is unknown. However, the results of the exercise were strengthened, as Lane et al. (2005) also consulted a group of experts who confirmed the items' relationship to BCC.

#### *3.3.3.1.2 Content and face validity*

Three articles examined content and face validity of the measures. Content validity refers to the degree to which items on a particular measure are relevant and representative of a given construct (Streiner et al., 2015). Whereas face validity denotes whether a measure appears (at face value) to measure what it is supposed to measure (Bowers et al., 2014). Both content and face validity differ from other forms of validity because they are not based on the item scores. Assessment of content and face validity takes place during scale development, usually through expert consultation and/or scale piloting (e.g., CTS-R; Blackburn et al., 2001a). Content validity can be quantified. For example, in developing the IAC Treatment Fidelity Instrument, Torrey (2012) asked two experts to rate the items on their clarity, relevance, and representativeness of the treatment constructs. The ratings were used to compute the content validity index (CVI), giving the proportion of experts who considered the items to be content valid (Rubio et al., 2003). Generally a CVI value of 0.78 or above is indicative of good content validity (Lynn, 1986); but in this article, the value was increased to 1.0 because only two experts were consulted (Torrey, 2012).

#### *3.3.3.1.3 Convergent validity*

Convergent validity of the measures was examined in six of the articles. This type of validity refers to the extent to which a particular measure is related to other measures that capture a common construct (Carlson and Herdman, 2012). Most articles assessed convergent validity using correlation analyses ( $n=5$ ); the relationships reported were generally in the expected directions, providing support of validity (e.g., YACS; Carroll et al., 2000). Pearson's correlation was the most commonly reported coefficient ( $n=4$ ). Pearson's correlation is appropriate for continuous variables that are approximately normally

distributed (Bowers et al., 2014). However, ordinal data is inherently non-normal (the identified measures were measured at the ordinal level); therefore, *“a test of the significance of Pearson's r may inflate Type I error rates and reduce power”*<sup>32</sup> (Bishara and Hittner, 2012, p.399). Deviations from the normality assumption may lead to false assumptions about the relationships, and obscure significant findings (Arndta et al., 1999).

It is argued that non-parametric correlation coefficients are more appropriate for ordinal data, or for continuous data that is non-normally distributed (Bowers, 2014). For example, Haddock et al. (2012) examined convergent validity of the MI-CTS using Spearman's correlation coefficient. The test makes no assumptions about the distribution of data, and is useful for minimising the effects of outliers (Field, 2017). However, Spearman's correlation generally underestimates the true (population) value, particularly when analysing small samples (Arndta et al., 1999). To avoid this bias, others maintain that parametric tests, like Pearson's correlation, are sufficiently robust, and *“can be used with Likert data, with small sample sizes, with unequal variances, and with non-normal distributions”* (Norman, 2010, p.631).

The sixth article used a different approach for examining convergent validity (Tober et al., 2008). Quality summary scores from the UKATT PRS were compared with scores from individual judges on the overall quality of therapists' treatment delivery. Scores were compared using analysis of variance (ANOVA). This simple approach also provided evidence of concurrent validity, as data collection occurred at the same time (Bowers et al., 2014). The reliability of the judges' quality assessment scores was not evidenced; low reliability can attenuate the validity estimates.

#### 3.3.3.1.4 Discriminant validity

Eleven articles examined discriminant validity of the measures. That is, the measures' ability to discriminate between different treatments. This type of validity was assessed by comparing mean scores from a particular measure across treatment groups. The most commonly used methods for making these comparisons were the two-sample t-test ( $n=3$ ), one-way ANOVA ( $n=3$ ), and contrast analyses ( $n=2$ ). The two-sample t-test describes the *“difference in means of two independent groups”* (Bowers et al., 2014, p.158). The method

---

<sup>32</sup> Type I errors (false-positive): rejection of the null hypothesis when it is actually true (Bowers, 2014).

is advantageous because it is generally robust to deviations from the distribution assumptions, although the approximation is less good for samples of unequal size (Bland, 2017); treatment groups were unequal in all three articles that used this method (e.g., ACS-SEC; Barber et al., 1997). Irrespective, violations of normality are unlikely to give “*spurious significant differences*”, rather the power of the test will be reduced (Bland, 2017, p.141). For large samples (e.g., UKATT PRS; Tober et al., 2008), the effect of using the test with non-normal data is minimised (Bowers et al., 2014; Bland, 2017).

A disadvantage of the t-test is multiple comparisons (e.g., ACS-SEC; Barber et al., 1997). The more groups that are compared, the more likely a significant difference will be found when the null hypothesis is true (Bland, 2017). One way to overcome this disadvantage is to compare several means (three or more groups) using ANOVA. This method provides information on whether the means are significantly different from one another, but not where that difference lies. It was, therefore, unsurprising that two (of the three) articles using ANOVA reported additional post hoc analyses (e.g., CSPRS-6; Hill et al., 1992). To reduce the likelihood of making a type I error, two articles used the Bonferroni correction ( $\alpha/n$  where  $n$  is the number of comparisons) for multiple comparisons (e.g., Hilsenroth et al., 2005; Kao and Green, 2008). Articles that conducted post hoc analyses also carried out a multiple-groups profile analysis. This method is akin to a multifactor repeated measures ANOVA approach (e.g., YACS; Carroll et al., 2000), and is not widely used in the health measurement literature. Contrast analyses were reported in two articles, the particular method used was not specified (e.g., ACS-IDCCD; Barber et al., 1996).

Other methods of examining treatment differentiation in the articles were: linear regression ( $n=1$ ), and the Chi-square test ( $n=1$ ):

- i) Linear regression is an extension of ANOVA. The method is preferred to ANOVA for more complex study designs, including unequal treatment groups (Bowers et al., 2014; Field, 2017). Linear regression was used by Watson et al. (2013a) to test for differences in AESOPS PRS subscale scores, separated for adherence and competence, across two treatment groups. Confidence intervals were provided, indicating the range of plausible values for the true population (du Prel et al., 2009). Evidence that the assumptions of linear regression (the least squares method) were examined was not provided. This requirement would have helped support the validity of the method (Bland,

2017). It is likely that the residuals followed non-constant variance, in which case a weighted least squares regression would have been more appropriate.

- ii) The Chi-square test was used by Pantaloni et al. (2012) to test for “*differences in the mean percentage occurrence of individual BAS items*” across treatment groups (p.386). The test applies to categorical data; items on the BAS were scored dichotomously (**Table 3**). The expected values for some items did not exceed one (e.g., item 9 ‘confront patient’; based on the observed values presented in the article), raising concerns about the validity of the test (Bland, 2017). In such cases, the conservative Fisher’s exact test is considered more appropriate (Bowers, 2014).

#### 3.3.3.1.5 Sensitivity to change

Sensitivity to change was examined in two articles. This type of validity describes the ability of a measure to detect changes in the attribute being assessed (Streiner et al., 2015).

Blackburn et al. (2001a) used paired t-tests to demonstrate improvements in therapist competence. The paired t-test compares the means of two matched or paired groups (Bowers et al., 2014). The article focused on trainee therapists who saw two clients at different stages of training. The results showed that total scores on the CTS-R for the second client were significantly higher than the total CTS-R scores for the first client. While significant differences were highlighted, paired t-tests do not provide information on the magnitude of the effect (Field, 2017).

Lane et al. (2005) used a similar research design, focusing on changes in BECCI scores before and after therapist training. However, sensitivity to change was examined using the standardised response mean (SRM), a measure of effect size. The SRM is advantageous because it provides an indication of the importance of the effect (Field, 2017). It is calculated by dividing the mean change (in a single group) by the standard deviation (SD) of change in scores (Streiner et al., 2015). The SRM was interpreted using guidelines provided by Cohen (Cohen, 1988): “*a score of 0.8 or above is thought to show a high level of responsiveness*” (Lane et al., 2005, p.169). These guidelines are based on the pooled SD of baseline scores, rather than the SD of change in scores (Middel and van Sonderen, 2002; Streiner et al., 2015). Applying Cohen’s guidelines for SRM values may, therefore, overestimate or underestimate the magnitude of effect (Middel and van Sonderen, 2002).



### **3.3.3.2 Reliability**

Reliability reflects the stability of a measure (Bowers et al., 2014), the degree to which measurements can be replicated (Koo and Li, 2016). The articles evaluated three types of reliability: internal consistency, inter-rater reliability (IRR), and test-retest reliability.

#### *3.3.3.2.1 Internal consistency*

Fourteen articles examined internal consistency of the measures. Internal consistency describes the extent to which item scores on a measure are correlated (Bowers et al., 2014). Acceptable levels of internal consistency were reported in all 14 articles (e.g., SPRS; Shapiro and Startup, 1992); although for seven measures, there was low reliability for some subscales (e.g., ACS-IDCCD; Barber et al., 1996). The most commonly reported coefficient for assessing internal consistency was Cronbach's alpha ( $n=13$ ). The coefficient denotes the extent to which items on a measure (or subscale) assess the same construct (Tavakol and Dennick, 2011).

Cronbach's alpha is dependent on both the inter-relatedness of the items, and the number of items on the measure (Streiner et al., 2015; Field, 2017). The value of alpha can be increased by adding more items, although this increase gets smaller as the measure becomes longer (Bland, 2017). For example, two articles reported alpha coefficients separately for each subscale and for the measure as a whole (e.g., ACS-IDCCD; Barber et al., 1996); unsurprisingly, the coefficients were larger for the entire measure than for each subscale. In one article, the values of alpha were very high, indicating some redundancy in the measure (CTS-R, Blackburn et al., 2001a; Bland, 2017).

The Kuder-Richardson coefficient was reported for only one measure, the BAS (Pantalon et al., 2012). The coefficient is similar to Cronbach's alpha, but is used only for measures that are scored dichotomously (Streiner et al., 2015), as was the case for the BAS.

#### *3.3.3.2.2 Inter-rater reliability*

IRR was the most commonly reported test of reliability, examined in all 21 articles. IRR refers to the consistency of measurement between two or more raters (Trevethan, 2017). Support for IRR of the measures was demonstrated (e.g., TSF-ACES; Campbell et al., 2013), although ten articles reported poor reliability for some of the scale and item scores (e.g., CTS-R; Blackburn et al., 2001a). Confidence intervals for the reliability estimates were presented in only two articles (AESOPS PRS, Watson et al., 2013a; IAC Treatment Fidelity

Instrument, Torrey, 2012), providing information on the magnitude of the effect (Bowers et al., 2014).

The most frequently reported statistic for examining IRR in the articles was the intraclass correlation coefficient (ICC) ( $n=19$ ). ICCs represents the proportion of variance that is attributable to 'true' differences between subjects, in this case the therapists (Streiner et al., 2015). There are different variants of the ICC (Shrout and Fleiss, 1979; McGraw and Wong, 1996) (Appendix B); information provided in the articles on the model (one-way random-effects, two-way random-effects, two-way mixed-effects), form (single measures, average measures), and type (consistency, absolute agreement) of ICC varied. Only ten articles detailed the full ICC variant used (e.g., ITRS; Martino et al., 2008). The remaining nine articles omitted information on the model, form, and/or type; for example, six articles did not specify the ICC model (e.g., CBT for PTSD Fidelity Scale; Lu et al., 2012).

Reporting the full ICC variant is important for assessing the appropriateness of the ICC variant for the study design, and for interpreting the IRR estimates (Hallgren, 2012); for example, mixed-effects models cannot be generalised to other raters, the results represent the reliability of the specific raters involved in the study (Koo and Li, 2016; Trevethan, 2017). The study conclusions in cases where the ICC variant was not specified should be interpreted with some caution. There is debate about using a parametric test, such as the ICC, to examine reliability. The ICC is primarily an index of reliability for continuous variables (Liu et al., 2016); however, items on the identified measures were scored at the ordinal level. An alternative reliability estimate for ordinal variables is the weighted kappa (Mandrekar, 2011).

Other tests of IRR used in the articles were: the Pearson's correlation coefficient ( $n=1$ ), percentage agreements ( $n=1$ ), and the generalisability coefficient ( $n=1$ ):

- i) Pearson's correlation coefficient (or Pearson's  $r$ ) provides an indication of the strength of association between two raters' scores (Bowers, 2014). The coefficient can be a valid estimator of IRR when there are meaningful pairings between two raters (Landers, 2015), as was the case in validating the MTRS (DeRubeis et al., 1982). The more flexible ICC is preferred to Pearson's correlation because: i) a single ICC can be calculated for two or more raters, and ii) the ICC will yield a value of 1.0 (complete reliability) when raters' scores

are identical, not when the scores for one rater are a linear function of the second (Streiner et al., 2015).

- ii) Percentage agreements were defined by Haddock et al. (2012) as “*the number of cases rated where there was complete agreement between the four raters (adherent or not)*” (p.42). Although simple and intuitive to use, percentage agreements do not account for chance agreement, and therefore overestimate the level of IRR (McHugh, 2012; Hallgren, 2012).
  
- iii) Generalisability coefficients represent the degree of association between the raters’ scores and the average scores of all possible raters. The test considers multiple sources of measurement error, and is estimated from ANOVA (Webb et al., 2006). Madson et al. (2005) reported the coefficient for two raters using MISTS, along with ICCs for the individual items. Generalisability coefficients provide a useful index of reliability. However, the associated confidence intervals are difficult to calculate (Streiner et al., 2015), and the test is not “*readily accessible to researchers because of its technical development and presentation*” (Alkharusi, 2012, p.194).

#### 3.3.3.2.3 Test-retest reliability

Test-retest was examined in only one article regarding the BECCI (Lane et al., 2005). This form of reliability refers to the degree of consistency of measurement within raters across time periods (Bowers et al., 2014). Evidence for test-retest reliability was provided using the ICC. The model and type of ICC were not specified. Incomplete information raises concerns about the correctness of the analyses, and makes comparisons between studies difficult (Koo and Li, 2016). Confidence intervals for the reliability estimates were not provided.

### 3.4 Discussion

#### 3.4.1 Study overview

A literature review was conducted to identify existing fidelity measures that evaluate the delivery of psychological therapies for addressing alcohol and drug use problems. The review was carried out in two stages using a systematic approach. The first stage identified articles from the literature that described the development and/or validation of fidelity

measures relevant to the addiction field. The second stage identified measures from the articles found in stage 1 that evaluated the delivery of psychological therapies used for alcohol and drug use problems. Characteristics of the measures that met the selection criteria were summarised. Consideration was also given to the methods used to develop and validate the identified measures; the main methods described in the articles were discussed.

### **3.4.2 Main findings**

#### **3.4.2.1 Characteristics of the identified measures**

Twenty-six fidelity measures were identified from the literature review (**Table 3**). Most were developed for use in the addiction field for research purposes. While some measures focused on specific mental health problems (e.g., depression), others adopted a broader approach focusing on a particular treatment modality (e.g., cognitive therapy). Some measures were designed for use in other contexts, including: training, supervision, and self-reflective practice. Typically the measures evaluated treatment adherence and/or therapist competence in delivering one particular modality, with CBT being the most common. There were 25 different therapies evaluated by the measures. Most of the therapies were considered to be widely used in the addiction field (e.g., motivational interviewing). While some therapies were not typically used in the field, they had been adapted for alcohol or drug use problems (e.g., supportive-expressive therapy).

The number of items on the measures ranged from 11 to 96; the mean was 33 items. Measures developed for sole use in research were generally longer (with a mean of 39 items) than those developed for other contexts (including research; with a mean of 19 items). Items were scored using four main response formats: i) adjectival scales – ordinal scales with unipolar response options, ii) Likert scales – ordinal scales with bipolar response options, iii) ordered-categorical scales – Likert-type scales with item-specific descriptive anchors, and iv) dichotomous scoring – binary response options. Adjectival scales were frequently used to rate treatment adherence, and Likert scales for rating therapist competence. The majority of measures scored items on 5-point and 7-point scales. In general, all points were labelled on the 5-point scales, and four points labelled on the 7-point scales. A ‘not applicable’ option was provided in some measures, typically for rating therapist competence when the specified behaviour was not observed during the session.

### 3.4.2.2 *Methods of validating the identified measures*

Twenty one articles examined the measures' psychometric properties. **Table 5** shows the main analytic methods used to evaluate the different types of reliability and validity. Inter-rater reliability and internal consistency were the most commonly evaluated properties, closely followed by tests of discriminant validity. A range of analytic methods were used to support reliability and validity of the measures. The methods described highlighted three main issues. The first issue focuses on the appropriateness of using parametric tests with ordinal data. Data obtained from the measures were often treated as continuous variables, enabling the administration of parametric tests, such as, the t-test and ICC (Kahler et al., 2008). Such procedures can lead to inaccurate conclusions, as the assumptions of normality and homogeneity of variance are untenable. Researchers have argued that parametric tests are robust to violations of assumptions (Norman, 2010; Mircioiu and Atkinson, 2017), particularly when sample sizes are large (Bland, 2017). However, it has been shown that these tests are only robust in certain circumstances; for example, with samples of equal size (Erceg-Hurn and Mirosevich, 2008; Bland, 2017). Even when a test is robust to Type I error, the power of the test will be reduced (Bland, 2017). In order to assess the appropriateness of the methods (in the context of scale validation), researchers should make explicit the decision making process for using parametric tests with ordinal data. The preference for parametric methods should not "*preclude the use of alternative procedures*" (Erceg-Hurn and Mirosevich, 2008, p.594).

The second issue relates to the size of the samples used in the articles. Tests with smaller samples may not be sufficiently powered to detect effects that "*are truly present in a population*" (Bowers et al., 2014, p.61). In such cases, researchers may fail to reject the null hypothesis, leading to a type II error. Only one article provided evidence of a power calculation<sup>33</sup> (IAC Treatment Fidelity Instrument; Torrey, 2012), indicating that the sample was sufficient in size to have reduced the chance of a false-negative conclusion (Noordzij et al., 2010). For the remaining articles, the uncertainty of detecting a true effect should have been discussed, ideally with reference to a confidence interval (CI) approach<sup>34</sup> (Altman et al., 2000; Bowers et al., 2014). Two of the articles did include CIs for the reliability and

---

<sup>33</sup> The calculation was based on the sample size needed for examining inter-rater reliability using the intraclass correlation coefficient; 49 recordings of therapy sessions were required.

<sup>34</sup> The width of the CI depends on the precision of the standard error, and hence both the variability of the outcome and the sample size (Altman et al., 2000). A larger sample will generally result in a narrower CI (du Prel et al., 2009).

**Table 5: Summary the main tests of reliability and validity**

Type of Reliability/Validity	Definition	Analytic Method
Internal consistency ( <i>n</i> =14)	The degree item scores on a measure are correlated (Bowers et al., 2014).	<ul style="list-style-type: none"> <li>• Chronbach's alpha (<i>n</i>=13)</li> <li>• Kuder-Richardson coefficient (<i>n</i>=1)</li> </ul>
Inter-rater reliability ( <i>n</i> =21)	The degree of consistency of measurement between two or more raters (Trevethan, 2017).	<ul style="list-style-type: none"> <li>• Intraclass correlation coefficient (<i>n</i>=19)</li> <li>• Pearson's correlation coefficient (<i>n</i>=1)</li> <li>• Percentage agreements (<i>n</i>=1)</li> <li>• Generalisability coefficient (<i>n</i>=1)</li> </ul>
Test-retest reliability ( <i>n</i> =1)	The degree of consistency of measurement within raters across time periods (Bowers et al., 2014).	<ul style="list-style-type: none"> <li>• Intraclass correlation coefficient (<i>n</i>=1)</li> </ul>
Construct validity ( <i>n</i> =9)	The ability of a measure to assess the underlying constructs it purports to measure (DeVellis, 2017).	<ul style="list-style-type: none"> <li>• Principal components analysis (<i>n</i>=3)</li> <li>• Exploratory factor analysis (<i>n</i>=2)</li> <li>• Confirmatory factor analysis (<i>n</i>=3)</li> <li>• Construct explication exercise (<i>n</i>=1)</li> <li>• Also evidenced by demonstrating content, convergent and discriminant validity.</li> </ul>
Content and face validity ( <i>n</i> =3)	Content validity is the degree to which items on a measure are relevant and representative of a given construct (Streiner et al., 2015). Face validity considers whether a measure appears (at face value) to measure what it is supposed to measure (Bowers et al., 2014).	<ul style="list-style-type: none"> <li>• Assessed during scale development through expert consultation and/or scale piloting (<i>n</i>=3).</li> <li>• Content validity index (<i>n</i>=1)</li> </ul>
Convergent validity ( <i>n</i> =6)	The extent to which a measure is related to other measures that capture a common construct (Carlson and Herdman, 2012).	<ul style="list-style-type: none"> <li>• Pearson's correlation coefficient (<i>n</i>=4)</li> <li>• Spearman's correlation coefficient (<i>n</i>=1)</li> <li>• Analysis of variance (<i>n</i>=1)</li> </ul>
Discriminant validity ( <i>n</i> =11)	The ability of a measure to discriminate between different treatments.	<ul style="list-style-type: none"> <li>• Two-sample t-tests (<i>n</i>=3)</li> <li>• Analysis of variance (<i>n</i>=3)</li> <li>• Multiple-groups profile analysis (<i>n</i>=2)</li> <li>• Contrast analyses (<i>n</i>=2)</li> <li>• Linear regression (<i>n</i>=1)</li> <li>• Chi-square test (<i>n</i>=1)</li> </ul>
Sensitivity to change ( <i>n</i> =2)	The ability of a measure to detect changes in the attribute being assessed (Streiner et al., 2015).	<ul style="list-style-type: none"> <li>• Paired t-tests (<i>n</i>=1)</li> <li>• Standardised response mean (<i>n</i>=1)</li> </ul>

Note. The bracketed numbers represent the number of articles that examined a particular type of reliability or validity/used a certain analytic method.

validity estimates (one of which was Torrey, 2012), providing an alternative means of evaluating the precision of the findings. Researchers should therefore, at the very least, include CIs in the write up of their results.

The third issue was the completeness of the article descriptions of the methods used to evaluate the psychometric properties of the measures. This issue was particularly relevant for the main tests of inter-rater reliability and construct validity. Inconsistent reporting practices were found for the ICC model, and the procedures for factor analysis. Incomplete information raises concerns about the correctness of the analyses, and makes comparisons between studies difficult (Koo and Li, 2016). When reporting the methods, researchers need to include enough information to enable readers to make an informed opinion about what was done to evaluate the measures' psychometric properties (Field, 2017).

### **3.4.3 Strengths and limitations**

This literature review provides a clear primary objective: to identify existing fidelity measures that evaluate delivery of psychological therapies for addressing alcohol and drug use problems. The scope was purposefully broad to capture the breadth of measures suitable for evaluating treatment delivery in the addiction field (Meline, 2006). Reviews addressing broader topics can lack the *"direction, clarity, and focus needed to inform subsequent stages of the research process"* (Levac et al., 2010, p.3). Therefore, explicit criteria were developed for the selection of articles and measures. Unlike systematic reviews, the criteria were refined during the review as knowledge of the topic increased (Arksey and O'Malley, 2005). Consistency was ensured by reappraising the relevance of the identified measures with the supervision team. While this study did not seek to assess the quality of included measures, consideration was given to the analytic methods used to examine the measures' psychometric properties. This consideration informed the psychometric work of the BATS.

A key strength of the review was that the search was thorough and comprehensive. Multiple sources were searched to identify as many relevant measures as possible, minimising the risk of selection bias (Lefebvre et al., 2008). For example, had only one database been searched, some articles would have been missed because the journals they are published in are not all listed in the one database (Bigby and Williams, 2003). Efforts were made to reduce publication bias by contacting the authors of relevant articles; five of the included measures were signposted by authors (i.e., the measures were not identified

from the literature). The risk of bias associated with missing data was not considered problematic. Measures excluded because they could not be obtained targeted similar therapies to those included in the review; for example, six of the excluded measures evaluated CBT.

A potential concern is that two reviewers did not independently assess the relevance of articles retrieved from the database and reference list searches (McDonagh et al., 2013). Using two reviewers can reduce the potential for relevant articles to be excluded (Higgins et al., 2003). However, the supervision team were involved throughout the selection process, resolving any uncertainty about whether an article or measure was eligible for inclusion in the review. In addition, the search process is transparent, adding confidence that the decisions made are reproducible.

### **3.5 Conclusion**

A literature review was conducted to identify fidelity measures that evaluate the delivery of psychological therapies used for alcohol and drug use problems. Twenty-six measures were identified. Most of the measures were developed for use in the addiction field for research purposes. In general, the measures evaluated treatment adherence and/or therapist competence in delivering one particular treatment modality, with cognitive-behavioural therapy being the most common. The number of items on the identified measures ranged from 11 to 96, with a mean of 33 items. The measures developed solely for use in research were generally longer than those developed for other contexts, such as, training and supervision. Adjectival scales were most frequently used to rate treatment adherence, and Likert scales to rate therapist competence. Generally, items were scored using 5-point and 7-point scales. A range of analytic methods were used to support reliability and validity of the measures. The measures identified from the literature formed the basis of generating items for the BATS. The next chapter details how the item pool for the BATS was generated.



## **Chapter 4**

### **Generating an item pool**

#### **4.1 Introduction**

The previous chapter described the first of four studies undertaken to develop the BATS. The first study was a literature review, which identified 26 fidelity measures that evaluate the delivery of psychological therapies for alcohol and drug use problems. These 26 measures formed the basis of the second study, which is described in this chapter. The second study aimed to generate items for potential inclusion in the BATS. This chapter details how the item pool was generated. Focus is also given to item scoring. The process of generating response formats for the BATS is presented.

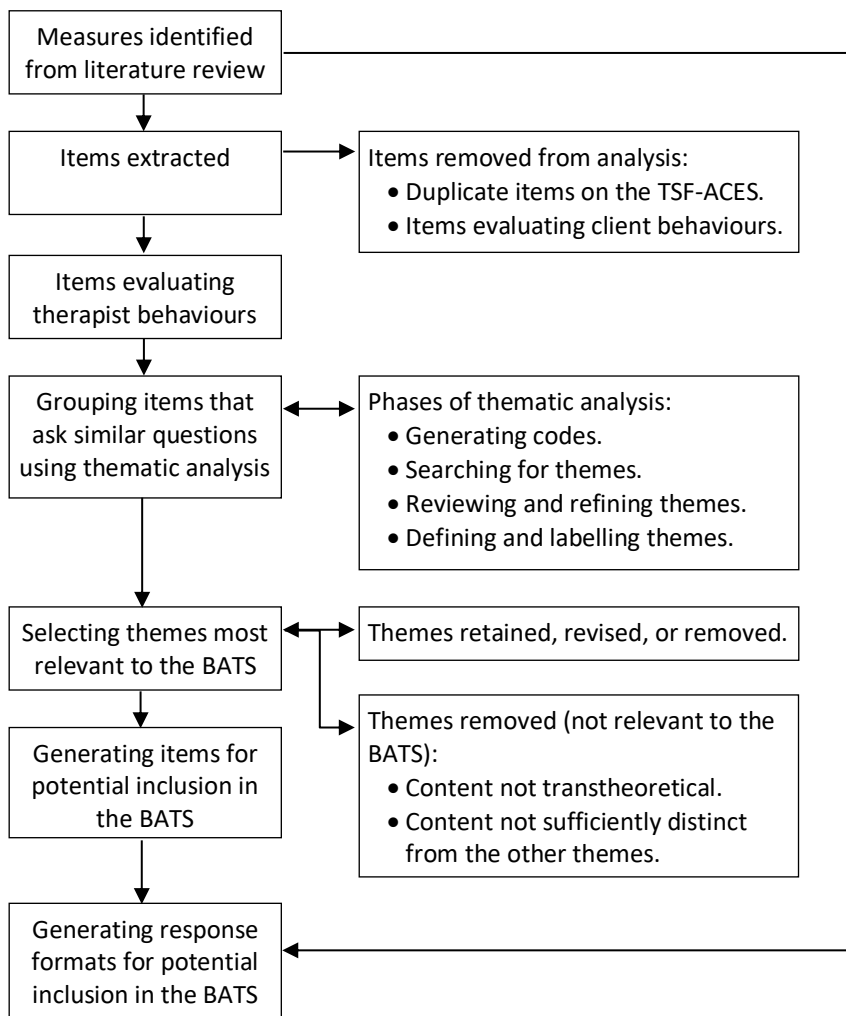
#### **4.2 Method**

Measures identified from the literature review in Study 1 were used as a basis for generating items for the BATS, and for informing a decision about how the items should be scored. Generation of the item pool was completed in three stages. First, items from the identified measures were analysed using a form of thematic analysis (Braun and Clarke, 2006); items were grouped based on what aspect of therapeutic practice they target. Second, the thematic structure (developed in stage 1) was refined, in collaboration with the supervision team (GL, BB, and GT), to identify the themes most relevant to the BATS; for example, themes concerning therapy specific techniques were removed as the BATS was to be transtheoretical. Third, themes identified as most relevant were expressed as individual items. These items were chosen to reflect that aspect of therapeutic practice. At this stage, consideration was also given to scoring. Generating response formats for the BATS was important for ensuring the items could be written in a consistent in format. An overview of the method is provided in **Figure 3**.

#### **4.3 Data**

All items were extracted from the identified measures with two exceptions. First, there

were duplicate items on the TSF-ACES<sup>35</sup> (Campbell and Guydish, 2012); 23 unique items were coded and the duplicate items were excluded from the analysis ( $n=34$ ). Second, items focusing on client behaviours (e.g., perceptions of the client’s motivation during the session) were excluded ( $n=50$ ). In total, 777 items were extracted from the 26 identified measures. The items described the activities of therapists, primarily the techniques associated with a particular treatment modality, such as, setting an agenda, and giving advice. Scoring methods used by the different measures were also considered.



**Figure 3: Methods for generating items and response formats for the BATS**

## 4.4 Data analysis

### 4.4.1 Stage 1: Grouping items that ask similar questions

The extracted items were analysed using a form of thematic analysis (TA); the purpose

<sup>35</sup> TSF-ACES = Twelve Step Facilitation Adherence Competence Empathy Scale

being to group items that asked similar questions. TA was chosen because it offers a flexible method for “*identifying, analysing and reporting patterns (themes) within data*” (Braun and Clarke, 2006, p.79). The method can be used for addressing different research questions and analysing different types of data (Clarke and Braun, 2013), and is useful for summarising key features of relatively large datasets (Nowell et al., 2017); data for this study included over 700 items. The form of TA used in this study was based on guidelines provided by Braun and Clarke (2006) and comprised six phases (**Table 6**).

**Table 6: Phases of the thematic analysis**

Phase	Overview of what the phase involved
1. Familiarisation with the data	The 26 identified measures were read and re-read. Initial analytic thoughts noted.
2. Generating codes	Items from the identified measures were coded ( $n=783$ ). This involved generating labels to identify salient features of the data.
3. Searching for themes	Codes were grouped together to form themes. Coded items were collated for each theme.
4. Reviewing and refining themes	Themes were reviewed and refined to better reflect the coded items, and the dataset as a whole.
5. Defining and labelling themes	Comprehensive definitions and names for each theme generated. Exemplar items, items that best reflected theme definitions selected.
6. Writing up	The analysis was written up to provide a coherent and plausible account of the data. The written narrative was supported by examples of the coded items.

The analytic process was recursive, proceeding to the next phase was not dependent on completing the phase prior (Clarke and Braun, 2013). The analysis was managed using NVivo11 for Windows. A more detailed overview of phases 2 to 5 are provided below.

#### **4.4.1.1 Generating codes**

Following familiarisation with the data, items from the identified measures were coded thematically. Coding involved generating labels, a word or short phrase, to identify salient features of the items (Clarke and Braun, 2013); that is, the aspect of therapeutic practice they were targeting. Initially, each composite part of an item was coded separately to capture the range of therapist behaviours evaluated (**Table 7**). This level of detail removed the coded extracts from their context (Joffe, 2012). As such, it was unclear whether each coded extract was a composite part or a complete item. The decision was, therefore, made to assign one code to each item (**Table 7**). Coding the main behaviours was advantageous

in developing a list of clearly defined codes that were distinct from one another (Joffe and Yardley, 2004). Generation of the codes was primarily inductive, “grounded in the content of the data”, some codes were more theoretically driven inspired by previous research in the area (Joffe, 2012, p.215).

#### 4.4.1.2 Searching for themes

The generated codes were used to search for themes. A theme was defined as “a coherent and meaningful pattern in the data” (Clarke and Braun, 2013, p.121). Codes of similar meanings (reflecting similar therapist behaviours) were grouped together. The relationships between the themes were considered; themes were combined to form meta-themes (higher level themes). Coded extracts were collated for each theme.

**Table 7: An example of how items were coded using multiple and exclusive codes**

Example Item	Coding Composite Parts	One Code per Item
<i>“Development of action plans</i>	Making a plan	Making a plan
<i>Identify realistic concerns</i>	Elicit client concerns	
<i>Establish feasible goal</i>	Treatment goals	
<i>Explore possible options</i>	Treatment options	
<i>Develop specific plan to address problem”</i>	Making a plan	

(CBT for PTSD Fidelity Scale; Lu et al., 2012)

**CBT for PTSD Fidelity Scale** = Cognitive-Behavioral Therapy for Post-Traumatic Stress Disorder Fidelity Scale.

#### 4.4.1.3 Reviewing and refining themes

Themes were refined to better reflect the coded extracts and the dataset as a whole (Clarke and Braun, 2013). Themes were refined in two ways: i) overlapping themes were combined, and ii) complex themes, in which several aspects could be identified, were split into sub-themes (lower level themes) (Table 8).

#### 4.4.1.4 Defining and labelling themes

The refined themes were labelled and defined. Comprehensive definitions were developed by referring back to the collated coded extracts for each theme. To do this, the coded items were organised into “a coherent and internally consistent account” (Braun and Clarke, 2006, p.92), ensuring that all aspects of the data were reflected in the theme definitions. The theme names and definitions were summarised in a table. Exemplar items

were included; these were coded items that best reflected the theme descriptions. Exemplars were purposively selected to reflect the range of identified measures. A thematic map was developed to visually present the themes and meta-themes.

**Table 8: Examples of how themes were refined**

Refinement	Themes: Definitions	Theme Refinements
Overlapping themes combined	<p>‘Signposting’: Therapists encouraging clients to contact and engage with self-help groups and other services.</p> <p>‘Case management’: Therapists’ facilitating and reviewing clients’ use of self-help groups and other services.</p>	<p>The content of the two themes overlapped; ‘signposting’ was combined with ‘case management’.</p> <p>The refined theme ‘case management’ was therefore defined as: Therapists encouraging, facilitating, monitoring and reviewing the clients’ use of self-help groups and other services.</p>
Complex themes split into sub-themes	<p>‘Empathic relationship’: Therapists communicating understanding of and sensitivity to the client’s perspective.</p>	<p>‘Empathic relationship’ was split into seven sub-themes to better reflect the different empathic behaviours.</p> <p>The refined sub-themes were labelled: i) empathy, ii) acceptance, respect, and being non-judgemental, iii) warmth and genuineness, iv) support and reassurance, v) being attuned and attentive, vi) rapport, and vii) sensitivity and concern.</p>

#### 4.4.2 Stage 2: Selecting themes most relevant to the BATS

The thematic structure developed in stage 1 was refined, in collaboration with the supervision team, to identify the themes most relevant to the BATS. Specifically, the researcher and the supervision team used the table, summarising theme names and descriptions, to make a decision on whether to retain, revise, or remove themes. The decision making process was distinct from stage 1; emphasis was given to selecting themes that reflected the key features of therapies widely used in the treatment of alcohol and drug use problems. The process was guided by three criteria: i) theme content was transtheoretical, ii) theme content was distinct from other themes – where multiple themes addressed a similar construct, the themes more conducive to transtheoretical working were prioritised, and iii) a clear exemplar was evident.

#### 4.4.3 Stage 3: Generating items for the BATS

Items for the BATS were based on the themes selected by the researcher and supervision team. Two approaches were used to generate the items. The first approach was to modify

an existing exemplar item. This was a time efficient approach, as the items had already been constructed (Streiner et al., 2015). The second approach was to construct a new item. These items were generally informed by several exemplar items. One item was generated for each selected theme. The items were listed in a table using the theme names as a reference. The items were generated in collaboration with the supervision team.

#### **4.4.3 Stage 4: Generating response formats for the BATS**

Possible response formats for the BATS were considered using the identified measures as a basis. A decision was made in collaboration with the supervision team on: i) the attributes being assessed (e.g., extensiveness, agreement, or skilfulness), ii) the scaling response (e.g., Likert scales, or dichotomous scoring), iii) the number of response options (e.g., 5-point scale), iv) the number of labelled anchors (e.g., end-anchored scales), and v) whether to include a 'not applicable' option.

### **4.5 Results**

#### **4.5.1 Stage 1: Grouping items that ask similar questions**

Items from the identified measures were sorted using thematic analysis, the purpose being to identify items that asked similar questions. Five meta-themes were developed (**Figure 4**). The first meta-theme, session management, focused on the techniques therapists may use to manage the therapy session; for example, setting an agenda, and explaining the purpose of the session. The second meta-theme, medication and case management, focused on therapists' discussion of the clients' medication and involvement in self-help groups and other services. The third meta-theme, interventions to increase awareness, concentrated on the techniques therapists may use to increase clients' understanding of their behaviours, thoughts, feelings and relationships; for example, exploring clients' conflicting thoughts about changing their behaviour. The fourth meta-theme, interventions to change behaviour, related to the techniques therapists may use to help clients change their behaviour and achieve their treatment goals; for example, giving advice, or encouraging clients to talk about behaviour change. The last meta-theme, core skills, related to how therapists delivered the session; for example, by asking questions, and developing an empathic and collaborative relationship.

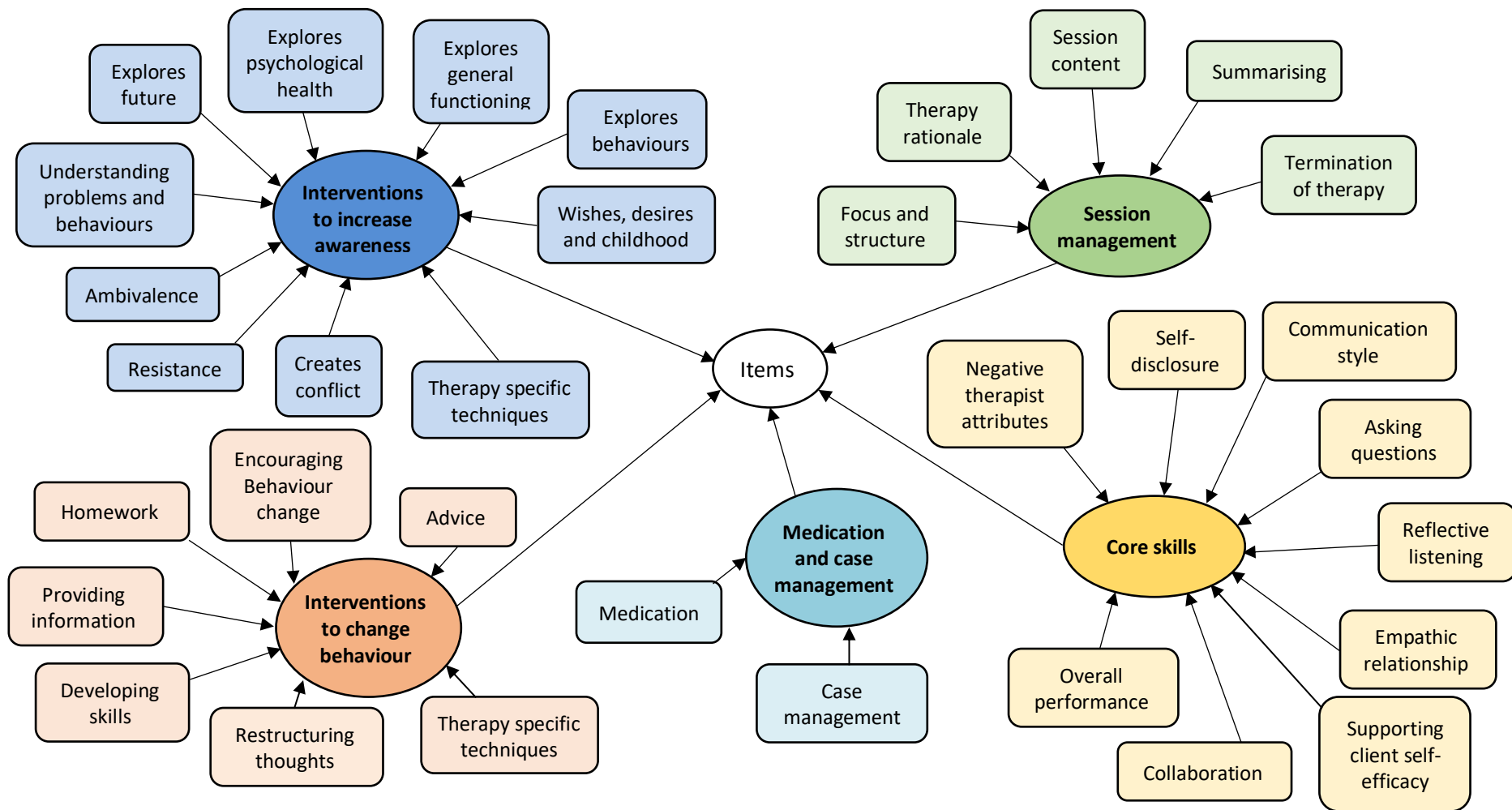


Figure 4: Thematic map showing the five meta-themes and associated themes

#### 4.5.1.1 Session management

The first meta-theme, session management, refers to therapists' overall management of the session (**Table 9**). There were five themes: i) focus and structure, ii) therapy rationale, iii) session content, iv) summarising, and v) therapy termination.

##### 4.5.1.1.1 Focus and structure

Focus and structure relates to how therapists structured the session. The theme captured five sub-themes: i) maintaining structure, ii) agenda setting, iii) consistency of problem focus, iv) directiveness, and v) time management.

- i) **Maintaining structure:** This was the main focus of seven items, taken from seven measures; it relates to therapists' attempts to maintain the session structure: *"To what extent did the therapist attempt to structure the session?"* (UKATT PRS; Middleton et al., 2001)<sup>36</sup>. While most items were relatively general ( $n=5$ ), two indicated how therapists might structure the session: *"Maintaining session structure (maintains session focus, sets appropriate tone and structure, appropriate level of CM clinician activity/directness, appropriate duration)"* (CM Clinician Rating Form; Petry and Stitzer, 2002)<sup>37</sup>.
- ii) **Agenda setting:** This was the salient feature of 14 items, taken from 13 measures. Most items were concerned with therapists setting and following an agenda ( $n=9$ ): *"To what extent did the therapist articulate and implement a specific agenda for the session (e.g., identify session topics, list issues to be discussed during the session)?"* (YACII; Nuro et al., 2005)<sup>38</sup>. MTRS<sup>39</sup> was the only scale to include two items, separating the setting and implementation of an agenda (DeRubeis et al., 1982). The remaining three items were more complex, combining agenda setting and collaboration: *"Did the therapist work collaboratively with the client to formulate and follow a specific agenda for the session?"* (SPRS; Shapiro and Startup, 1990)<sup>40</sup>.

---

<sup>36</sup> UKATT PRS = UK Alcohol Treatment Trial Process Rating Scale

<sup>37</sup> CM Clinician Rating Form = Contingency Management Clinician Adherence/Competence Rating Form

<sup>38</sup> YACSI = Yale Adherence and Competence Scale Second Edition.

<sup>39</sup> MTRS = Minnesota Therapy Rating Scale.

<sup>40</sup> SPRS = Sheffield Psychotherapy Rating Scale.



- iii) **Consistency of problem focus:** This was the main focus for five items, coded from five measures; it refers to therapists keeping the session focused on relevant topics: *“To what extent did you attempt to keep the session focused on prescribed activities (e.g., by redirecting dialogue when it strayed off tasks, by organizing the session so defined tasks were covered)?”* (CBT Therapist Checklist, Carroll, 1997)<sup>41</sup>. Items were similarly written, excepting one item taken from CTACS<sup>42</sup>. The CTACS item included item-specific descriptive anchors; for example, a high quality rating was achieved if *“the therapist used time extremely effectively by directing the flow of conversation and redirecting when necessary. Session seemed well-paced, focused, and structured”* (CTACS; Liese et al., 1995).
- iv) **Directiveness:** This was the salient feature for 10 items, coded from six measures; it relates to the level of direction or guidance given by therapists during the session. There was a distinction between therapists initiating what was discussed during the session ( $n=4$ ), and allowing the client to *“initiate the discussion of significant issues, events, and experiences”* (CPPS; Hilsenroth et al., 2005)<sup>43</sup> ( $n=4$ ). Other items focused on both therapist and client ( $n=2$ ): *“In general, the person who initiated changes in the flow of the direction of the session was the 1-therapist 9-client”* (MTRS; DeRubeis et al., 1982).
- v) **Time management:** This was the key feature for four items, taken from four measures; it refers to therapists’ use of time during the session: *“Did the therapist pace the session appropriately?”* (MACT Rating Scale; Davidson et al., 2004)<sup>44</sup>. Items varied in complexity, from a brief sentence, *“pacing/use of time”* (MI-CTS; Haddock et al., 2012)<sup>45</sup>, to a detailed description of what constituted *“effective time management”* (CTS-R; Blackburn et al., 2001b)<sup>46</sup>.

#### 4.5.1.1.2 Therapy rationale

Therapy rationale was the main focus for 18 items, coded from 10 measures; it relates to

---

<sup>41</sup> CBT Therapist Checklist = Cognitive Behavioral Therapy Therapist Checklist.

<sup>42</sup> CTACS = Cognitive Therapy Adherence and Competence Scale.

<sup>43</sup> CPPS = Comparative Psychotherapy Process Scale.

<sup>44</sup> MACT Rating Scale = Manual Assisted Cognitive Therapy Rating Scale.

<sup>45</sup> MI-CTS = Integrated motivational Interviewing and cognitive-behaviour therapy fidelity scale.

<sup>46</sup> CTS-R = Cognitive Therapy Scale – Revised.

therapists' explanation of the overall purpose of the therapeutic approach and/or the reasons for pursuing a particular topic in the session. Eight items were transtheoretical; a specific therapy was not specified: *"The therapist explains the rationale behind his or her technique or approach to treatment"* (CPPS; Hilsenroth et al., 2005). The remaining 10 items focused on a particular treatment approach, such as, interpersonal therapy, and twelve-step recovery:

*"To what extent did the therapist explicitly refer to Twelve Step Recovery or interpret or explain a particular Step to the patient or invoke a particular Step during the session or discuss the client's progress through the Steps?"* (YACSII; Nuro et al., 2005)

#### 4.5.1.1.3 Session content

Session content encapsulated 24 items, taken from 11 measures; it relates to both the therapeutic approach and the behavioural techniques delivered by therapists during the session. Six items focused on the therapeutic approach: *"Is this a Cognitive session? (Y/N) How confident are you of your answer?"* (ACS-IDCCD; Mercer et al., 1995)<sup>47</sup>. Seven items were concerned with the delivery of prescribed behaviours; for example, three measures included session content checklists (e.g., UKATT PRS; Middleton et al., 2001). The remaining 11 items focused on proscribed behaviours: *"To what extent did the therapist engage in family or relationship therapy?"* (CE Therapist Rating Form, Carroll et al., 1999)<sup>48</sup>.

#### 4.5.1.1.4 Summarising

Summarising was the main focus for 17 items, coded from 13 measures; it refers to therapists' use of summaries during the session. Items focused on therapists summarising the content of discussion from a previous session ( $n=6$ ), in the current session ( $n=3$ ), and at the end of the session ( $n=5$ ): *"To what extent did the therapist summarise at the end of the session, the content of the discussion that occurred during the session?"* (UKATT PRS; Middleton et al., 2001). The remaining three items were more complex, combining therapists' use of summaries from a previous session and the current session: *"Did the therapist summarize or encourage the client to summarize key issues discussed either in a*

---

<sup>47</sup> ACS-IDCCD = Adherence/Competence Scale for Individual Drug Counseling for Cocaine Dependence.

<sup>48</sup> CE Therapist Rating Form = Compliance Enhancement Therapist Adherence/Competence Rating Form.

**Table 9: An overview of ‘session management’ (meta-theme 1) with theme descriptions and exemplar items**

Theme	Sub-theme	Description	Exemplar(s)
Focus and structure	Maintaining structure	Items focused on whether therapists maintained an overall session structure.	<i>“To what extent did the therapist attempt to structure the session?”</i> (UKATT PRS; Middleton et al., 2001)
	Agenda setting	Items looked at whether therapists set and followed an agenda during a session.	<i>“To what extent did the therapist articulate and implement a specific agenda for the session (e.g., identify session topics, list issues to be discussed during the session)?”</i> (YACSII; Nuro et al., 2005)
	Consistency of problem focus	Items related to whether therapists kept the session focused on relevant target problems.	<i>“To what extent did the practitioner attempt to keep the session focused on target problems (e.g. redirecting dialogue when it strayed off tasks, by organizing the session so defined tasks were covered)?”</i> (AESOPS PRS; Tober and Crosby, 2011)
	Therapist directiveness	Items focused on the extent to which therapists directed the flow of conversation during the session.	<i>“In general, the person who initiated changes in the flow of the direction of the session was the: 1-therapist 9-client”</i> (MTRS; DeRubeis et al., 1982).
	Time management	Items referred to therapists’ use of time during the session.	<i>“Pacing and Efficient Use of Time Session length kept to 1 hour, Efficient structuring of time, Tactful limiting of peripheral &amp; unproductive discussion.”</i> (CBT for PTSD Fidelity Scale; Lu et al., 2012)
Therapy rationale		Items referred to therapists’ explanation of the therapeutic approach and/or the reasons for pursuing a particular topic in a session.	<i>“To what extent did the therapist explain the overall purpose of the particular treatment approach or the particular treatment session which is about to occur?”</i> (UKATT PRS; Middleton et al., 2001)
Session content		Items focused on the type of therapy or techniques delivered by therapists during the session.	<i>“Is this a Cognitive session? (Y/N) How confident are you of your answer?”</i> (ACS-SEC; Barber, 1997)

Theme	Sub-theme	Description	Exemplar(s)
Summarising		Items focused on whether therapists' summarised what was said: during the session, at the end of the session, and/or in previous sessions.	<i>"Did the therapist end the session by summarizing the main points covered? 1-did not 9-extensive and detailed"</i> (MTRS; DeRubeis et al., 1982)
Termination of therapy		Items concerned therapists discussing the termination of therapy.	<i>"To what extent did you discuss the termination of the therapy (e.g., encourage the patient to discuss his/her feelings or thoughts about termination, discuss plans for the continuation of treatment after the end of the study protocol)?"</i> (CBT Therapist Checklist, Carroll, 1997).

**ACS-SEC** = Adherence/Competence Scale for Supportive-Expressive Therapy for Cocaine Dependence; **AESOPS PRS** = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale; **CBT for PTSD Fidelity Scale** = Cognitive-Behavioral Therapy for Post-Traumatic Stress Disorder Fidelity Scale; **CBT Therapist Checklist** = Cognitive Behavioral Therapy Therapist Checklist; **MTRS** = Minnesota Therapy Rating Scale; **UKATT PRS** = UK Alcohol Treatment Trial Process Rating Scale; **YACSII** = Yale Adherence & Competence Scale Second Edition.

*previous session or in the current session” (CSPRS-6; SPR Project Staff, 1984)<sup>49</sup>*

#### **4.5.1.1.5 Termination of therapy**

Termination of therapy was the main topic for four items, taken from four measures. The items were complex, covering a range of behaviours; for example, exploring clients’ thoughts and feelings about ending therapy:

*“To what extent did the therapist make attempts to prepare the patient for the end of the patient/therapist relationship/treatment by exploring the patient’s thoughts/feelings regarding the termination of the relationship and/or preparing the patient for the termination of the treatment?” (YACSII; Nuro et al., 2005)*

#### **4.5.1.2 Medication and case management**

The second meta-theme, medication and case management, refers to therapists’ discussion of clients’ medication, and involvement in self-help groups and other services (**Table 10**). There were two themes: i) medication, and ii) case management.

##### **4.5.1.2.1 Medication**

Medication was the main focus for 34 items, taken from four measures; it relates to therapists’ assessment and review of clients’ medication. Most items concentrated on issues to be addressed prior to taking medication ( $n=21$ ); for example, possible effects, both positive and negative ( $n=5$ ), and client and family concerns ( $n=4$ ): *“To what extent did the therapist address the patient’s concerns about medication?”* (CE Therapist Rating Form; Carroll et al., 1999). Twelve items focused on therapists’ review of the clients’ medication:

*“To what extent did the therapist inquire about or discuss the patient’s compliance/ noncompliance with the prescribed study medication regime since the last session?” (YACSII; Nuro et al., 2005)*

Four items were concerned with therapists making a connection between clients’ medication and any behaviour change, including occurrence of side effects: *“Did the therapist relate positive change (or lack of change) to the treatment medication the client was receiving?”* (CSPRS-6; SPR Project Staff, 1984). The remaining item combined therapists’ assessment and review of clients’ medication:

---

<sup>49</sup> CSPRS-6 = Collaborative Study Psychotherapy Rating Scale – Form 6.

**Table 10: An overview of ‘medication and case management’ (meta-theme 2) with theme descriptions and exemplar items**

Theme	Subtheme	Description	Exemplar(s)
Medication		Items were concerned with therapists discussing clients’ medication, e.g., addressing client concerns, and assessing compliance.	<p><i>“To what extent did the therapist assess the patient’s concerns about taking study medication and address those concerns?” (YACSII; Nuro et al., 2005)</i></p> <p><i>“To what extent did the therapist assess medication compliance since the last session?” (CE Therapist Rating Form; Carroll et al., 1999)</i></p> <p><i>“Did the therapist relate positive change (or lack of change) to the treatment medication the client was receiving?” (CSPRS-6; SPR Project Staff, 1984)</i></p>
Case management		Items related to therapists encouraging, facilitating, monitoring and reviewing the clients’ use of self-help groups (e.g., alcoholics anonymous) and other services (e.g., primary care services and housing).	<p><i>“To what extent did the clinician discuss or facilitate the coordination of additional services (i.e., ancillary or adjunctive to primary substance abuse counseling), including those that might be provided by the clinic (e.g., psychiatric appointment, childcare, parenting groups) or other agencies (e.g., housing, vocational, educational, legal, medical, domestic violence services, financial/ insurance/ entitlements, transportation)? To what extent was the importance of these extra services emphasized, forms/releases filled out, appointments scheduled, or phone calls planned?” (ITRS; Martino et al., 2009b)</i></p>

**CE Therapist Rating Form** = Compliance Enhancement Therapist Adherence/Competence Rating Form; **CSPRS-6** = Collaborative Study Psychotherapy Rating Scale – Form 6; **ITRS** = Independent Tape Rater Scale; **YACSII** = Yale Adherence and Competence Scale Second Edition

*“To what extent did the clinician facilitate a discussion about the client’s medical problems that complicate his/her substance abuse treatment? To what extent did the clinician discuss or review medications for the treatment of medical, substance abuse, or psychiatric problems?” (ITRS; Martino et al., 2009b)<sup>50</sup>*

#### 4.5.1.2.2 Case management

Case management was the main focus for 16 items, taken from eight measures; it relates to therapists’ encouragement, facilitation, and review of clients’ use of self-help groups (e.g., alcoholics anonymous) and other services (i.e., primary care services, and housing). Most items were concerned with self-help group participation (n=13): *“To what extent did the therapist make a specific referral to a self-help group?”* (CE Therapist Rating Form; Carroll et al., 1999). Two of these items were multifaceted:

*“To what extent did the clinician encourage, monitor, or reinforce the client’s involvement in 12 Step (AA/NA/CA) or other recovery self-help meetings (e.g., relying on members, planning or participating in meeting-related activities)? To what extent did the clinician explicitly refer to or explain the principles (e.g., specific Steps or recovery concepts) or review the client’s progress in self-help groups?”<sup>51</sup> (ITRS; Martino et al., 2009b)*

The remaining two items focused on clients’ use of services: *“Encourage patient to follow-up with his/her Primary Care Practitioner (PCP)?”* (BAS; Pantalon et al., 2012)<sup>52</sup>.

#### 4.5.1.3 Interventions to increase awareness

The third meta-theme, interventions to increase awareness, refers to the techniques therapists may use to increase clients’ understanding of their problems, thoughts, feelings, and relationships (**Table 11**); these ranged from relatively simple assessment techniques to more sophisticated reflective listening skills. There were 10 themes: i) explores behaviours, ii) explores psychological health, iii) explores future, iv) explores general functioning, v) ambivalence, vi) creates conflict, vii) understanding problems and behaviours, viii) resistance, ix) wishes, desires and childhood, x) therapy specific techniques.

##### 4.5.1.3.1 Explores behaviours

Explores behaviours relates to therapists’ exploration of clients’ problem behaviours. The

---

<sup>50</sup> ITRS = Independent Tape Rater Scale

<sup>51</sup> AA = Alcoholics Anonymous; NA = Narcotics Anonymous; CA = Cocaine Anonymous.

<sup>52</sup> BAS = Brief Negotiation Interview Adherence Scale

theme comprised two sub-themes:

- i) **Explores general behaviours:** This was the main focus for four items, taken from four measures; it refers to therapists exploring clients' problem behaviours without reference to substance use and other concerns: "*Therapist focuses on client's current, identifiable, problematic behaviour*" (TPRS, Fisher et al., 2000)<sup>53</sup>.
- ii) **Explores substance use behaviours:** This was the main topic for 29 items, coded from 11 measures; it relates to therapists' assessment and monitoring of clients' substance use. Most items focused on issues associated with substance use problems; for example, patterns of use, triggers and cravings, and results of recent urine/breath tests ( $n=17$ ): "*If the patient self-reported substance use, to what extent did the CM clinician relate self-report of substance use to objective indicators of substance use?*" (CM Clinician Rating Form; Petry and Stitzer, 2002). Some items distinguished between clients' past and recent use of alcohol or drugs ( $n=10$ ): "*To what extent did the therapist establish a history of or characterize the patient's current episode of drug use?*" (CE Therapist Rating Form; Carroll et al., 1999). The remaining two items were concerned with therapists eliciting client concerns about their drinking: "*To what extent did the therapist attempt to elicit concerns from the client about drinking?*" (UKATT PRS; Middleton et al., 2001).

#### 4.5.1.3.2 Explores psychological health

Explores psychological health was a salient feature for eight items, taken from five measures; it refers to therapists' assessment of clients' psychological health. Most items were concerned with clients' psychiatric history and/or current symptoms ( $n=6$ ): "*To what extent did the therapist establish a history of or characterize the patient's current episode of psychiatric symptoms/problems?*" (CE Therapist Rating Form; Carroll et al., 1999). Two items were more specific, concentrating on depressive symptoms and suicide ideation: "*How completely did the therapist assess the client's suicide ideation and potential?*" (CSPRS-6; SPR Project Staff, 1984).

---

<sup>53</sup> TPRS = Therapy Process Rating Scale



#### 4.5.1.3.3 Explores general functioning

Explores general functioning was the main focus for six items, taken from six measures; it concerns therapists' assessment of clients' general level of functioning. Most items were similarly written ( $n=5$ ): *"Did the therapist survey the client's general level of functioning in major life spheres (e.g., work, intimate relationships, family life, social life, etc.)?"* (CSPRS-6; SPR Project Staff, 1984). The remaining item focused on two aspects of general functioning: *"current employment"*, and *"money management"* (ACS-IDCCD; Mercer et al., 1995).

#### 4.5.1.3.4 Explores future

Explores future was the main focus for one item: *"The therapist focuses on the patient's future life situations"* (CPPS; Hilsenroth et al., 2005).

#### 4.5.1.3.5 Ambivalence

Ambivalence was the key topic for six items, taken from six measures; it refers to therapists' exploration of clients' conflicting thoughts and feelings about changing behaviour. Three items were relatively simple: *"To what extent did you attempt to focus on the patient's ambivalence about changing their level of cocaine use?"* (CBT Therapist Checklist; Carroll, 1997). The remaining three items were multifaceted, indicating how ambivalence might be explored:

*"To what extent did the therapist address or explore the positive and negative effects or results of the client's substance use and what might be gained and lost by abstinence or reduction in substance use? To what extent did the therapist use decisional balancing, complete a cost-benefits analysis, or develop a list of pros and cons of substance use? How much did the therapist express appreciation for ambivalence as a normal part of the change process?"* (YACSII, Nuro et al., 2005)

#### 4.5.1.3.6 Creates conflict

Creates conflict was the salient feature for five items, taken from five measures; it concerns therapists' attempts to heighten discrepancies experienced by clients within conflicting aspects of their life. Most items concerned the conflict between where clients currently are and where they want to be ( $n=4$ ):

*"To what extent did the therapist create or heighten the internal conflicts of the client relative to his/her substance use? To what extent did the therapist*

*facilitate or increase the client's awareness of a discrepancy between where his/her life is currently versus where s/he wants to be in the future? How much did the therapist explore the role of substances in preventing the client from reaching life goals or values?" (YACSII; Nuro et al., 2005)*

Dependency conflicts were addressed by the remaining item: *"Did the therapist focus on conflicts between the client's desire for independence and for dependence as an important factor in the client's problems?" (CSPRS-6; SPR Project Staff, 1984).*

#### *4.5.1.3.7 Understanding problems and behaviours*

Understanding problems and behaviours refers to therapists helping clients to gain an understanding of their problems and behaviours. Items had an emphasis on increasing insight, rather than assessment or monitoring (which may indirectly help clients gain a better understanding). There were 4 sub-themes: i) understanding behaviours, ii) understanding relationships, iii) understanding thoughts and feelings, and iv) therapy specific techniques for increasing client awareness.

- i) ***Understanding behaviours:*** This was a relatively large sub-theme encompassing 44 items, coded from 18 measures; it relates to therapists helping clients to gain an understanding of their problem behaviours. Sixteen items were concerned with therapists making connections between clients' substance use and other problems (e.g., negative consequences, physical and mental health): *"To what extent did the practitioner attempt to elicit concerns from the client about drinking?" (AESOPS PRS; Tober and Crosby, 2011)<sup>54</sup>.*

Other items sought to understand clients' problem behaviours in relation to their thoughts, feelings, and relationships. The majority of these items were relatively simple, covering one particular issue, e.g., relationships ( $n=21$ ): *"To what extent did the therapist explore how drug use or craving since the last session related to specific interpersonal problems that were taking place at the time of the craving or slip?" (YACSII; Nuro et al., 2005).* Other items were more complex, including item-specific descriptive anchors and/or covering a range of issues ( $n=7$ ): *"The therapist should help the patient to gain an understanding of*

---

<sup>54</sup> AESOPS PRS = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale

*how his/her perceptions and interpretations, beliefs, attitudes and rules relate to his/her problem"* (CTS-R; Blackburn et al., 2001b).

- ii) ***Understanding relationships***: This was the main focus for 48 items, taken from eight measures; it relates to therapists helping clients' to understand their interpersonal relationships. Some items focused on clients' relationships more generally ( $n=6$ ): *"To what extent did the therapist focus on the client's interpersonal relationships and role expectations?"* (CSPRS-6; SPR Project Staff, 1984). Other items were more specific ( $n=13$ ); for example, nine items explored the similarities among clients' past and present relationships.

Emphasis was given to the negative and positive aspects of the clients' relationships. Relationship difficulties, including bereavement, were the main focus for 13 items: *"The therapist clarified and interpreted facets of the main relationship problem and conflict, i.e. provided explicit feedback to the patient about his/her interpersonal problems."* (ACS-SEC; Barber et al., 1997)<sup>55</sup>. Four items concentrated on the positive aspects: *"Did the therapist help the client to explore the best aspects of the client's prior relationships as a means of providing a model for the development of satisfying new relationships?"* (CSPRS-6; SPR Project Staff, 1984).

Some items related to therapists helping the client to work through their relationship difficulties ( $n=5$ ): *"Did the therapist help the client to consider ways in which the client can bring about desired changes in her/his interpersonal relationships (or role expectations in those relationships)?"* (CSPRS-6; SPR Project Staff, 1984). Role change was an associated concept, and was addressed by three items: *"To what extent did the therapist focus on the patient's current transition from one social role to another?"* (YACSII; Nuro et al., 2005).

The remaining four items focused on therapists' exploring the therapeutic relationship; for example, as a means of understanding the client's

---

<sup>55</sup> ACS-SEC = Adherence/Competence Scale for Supportive-Expressive Therapy for Cocaine Dependence.

interpersonal relationships: *“Did the therapist attempt to use the therapeutic relationship to help the client understand how she/he relates to others OR as a model for the client to use in developing satisfying relationships outside of therapy?”* (CSPRS-6; SPR Project Staff, 1984).

- iii) ***Understanding thoughts and feelings***: This was the largest sub-theme and included 54 items, taken from 11 measures; it refers to therapists increasing the clients’ awareness of their thoughts and feelings. Some items focused on clients’ being able to experience and express feelings ( $n=7$ ): *“The therapist encourages the patient to experience and express feelings in the session”* (CPPS; Hilsenroth et al., 2005). Three of these items focused on the client expressing feelings appropriately; *“Emotional levels that are too high or too low are likely to interfere with therapy”* (CTS-R; Blackburn et al., 2001b).

Other items targeted specific feelings, including those of which the client was unaware ( $n=3$ ), or found painful or uncomfortable to experience ( $n=4$ ). Some focused on reducing client distress: *“The therapist changed topics or otherwise directed the client away from painful or emotionally charged material.”* (TPRS; Fisher et al., 2000). Six items addressed clients’ thoughts and feelings more generally: *“Practitioner asks questions to elicit how patient thinks and feels about the topic”* (BECCI; Lane, 2002)<sup>56</sup>.

The remaining majority of items focused on increasing clients’ awareness of their thoughts and beliefs ( $n=30$ ): *“did the therapist examine (or encourage the client to examine) the validity of the client’s beliefs?”* (MTRS; DeRubeis et al., 1982). Eleven of these items were taken from CSPRS-6 (SPR Project Staff, 1984); the same items were also included in SPRS (Shapiro and Startup, 1990). CSPRS-6 and SPRS items were more specific than those from the other measures: *“Did the therapist help the client to use currently available evidence or information (including the client’s prior experiences) to test the validity of the client’s beliefs?”* (CSPRS-6; SPR Project Staff, 1984).

---

<sup>56</sup> BECCI = Behaviour Change Counselling Index.

#### 4.5.1.3.8 Resistance

Resistance or reactance was the main focus for eight items, taken from seven measures; it refers to therapists discussing clients' resistance to change their behaviour and engage in treatment. Most items were concerned with therapists exploring clients' resistance ( $n=5$ ); three of these items provided examples of client resistant behaviours: *"To what extent did the therapist explore the patient's denial/resistance to Twelve Step recovery (e.g., avoiding meetings, minimizing negative consequences..."* (YACSII; Nuro et al., 2005). One item concentrated on problem-solving client resistance to attend self-help groups: *"NA/CA/AA"* (CE Therapist Rating Form; Carroll et al., 1999). The remaining two items looked at how therapists reacted to client resistance; for example, in MISTS Revised<sup>57</sup>, a low competence rating was given to the therapist who *"argues with the client in favor of change"* (Madson and Loignon, 2007).

#### 4.5.1.3.9 Wishes, desires, and childhood

This was a key feature for nine items, taken from five measures; it relates to clients' wishes, dreams, and/or childhood experiences. Four items focused on therapist exploration: *"Therapist tries to uncover early experiences and unconscious wishes as a way of producing insight"* (TPRS; Fisher et al., 2000). The remaining items were concerned with therapists' making connections between clients' current problems (e.g., substance use) and their wishes and desires ( $n=2$ ), and childhood experiences ( $n=3$ ): *"Did the therapist suggest or imply that the client's current problems are related to a conflict or situation from her/his childhood?"* (CSPRS-6; SPR Project Staff, 1984).

#### 4.5.1.3.10 Therapy specific techniques for increasing awareness

This theme encapsulated 35 items, coded from nine measures; it refers to therapy specific techniques for increasing clients' awareness of their problems and behaviours. Because the BATS was designed to be transtheoretical, a separate theme was developed for therapy specific items not easily categorised into the aforementioned themes; for example, three items focused on guided discovery (e.g., CTACS; Liese et al., 1995), a process used in cognitive behavioural therapy (CBT) to help clients uncover, examine, and test their thoughts and experiences (Neenan and Dryden, 2006).

While some items were relatively simple ( $n=5$ ): *"Did the therapist focus on what she/he,*

---

<sup>57</sup> MISTS Revised = Motivational Interviewing Supervision and Training Scale Revised.

**Table 11: An overview of ‘interventions to increase awareness’ (meta-theme 3) with theme descriptions and exemplar items**

Theme	Sub-theme	Description	Exemplar(s)
Explores behaviours	Explores general behaviours	Items focused on therapists exploring clients’ problem behaviours, without referencing specific concerns.	<i>“Practitioner encourages patient to talk about current behaviour or status quo”</i> (BECCL; Lane, 2002)
	Explores substance use behaviours	Items related to therapists’ assessment of clients’ substance use problems, e.g., past and recent use, triggers and cravings, and results of recent urine/breath tests.	<i>“To what extent did the clinician maintain focus during the session on the client’s past or recent use of drugs and alcohol, including the pattern of use, extent of urges/thoughts, extent of reduction in use, results of recent urine/breath tests?”</i> (ITRS; Martino et al., 2009b)
Explores psychological health		Items related to therapists’ assessment of the client’s psychological health, e.g., psychiatric history and current symptoms.	<i>“To what extent did the clinician explicitly focus on the client’s psychopathology (i.e., symptoms of depressive, anxiety, psychotic disorders)? How much did the clinician discuss the client’s past and current psychiatric symptoms or treatment for a psychiatric disorder?”</i> (ITRS; Martino et al., 2009b)
Explores general functioning		Items were concerned with therapists’ assessment of the clients’ general level of functioning, e.g., relationships, and work.	<i>“To what extent did you assess the patient’s general level of functioning in major life spheres (e.g., work, intimate relationships, family life, social life, etc.)?”</i> (CBT Therapist Checklist; Carroll, 1997)
Explores the future		Item focused on therapists addressing the clients’ future.	<i>“The therapist focuses on the patient’s future life situations.”</i> (CPPS; Hilsenroth et al., 2005)
Ambivalence		Items focus on whether therapists explored clients’ ambivalence, conflicting thoughts and feelings, about changing their behaviour.	<i>“To what extent did the therapist address or explore the positive and negative effects or results of the client’s substance use and what might be gained and lost by abstinence or reduction in substance use? To what extent did the therapist use decisional balancing, complete a cost-benefits analysis, or develop a list of pros and cons of substance use? How much did the therapist express appreciation for ambivalence as a normal part of the change process?”</i> (YACSII, Nuro et al., 2005)

Theme	Sub-theme	Description	Exemplar(s)
Creates conflict		Items related to therapists heightening discrepancies experienced by clients on conflicting aspects of their life.	<i>"To what extent did the therapist create or heighten the internal conflict or discrepancy experienced by the client between where they are currently and where they want to be."</i> (UKATT PRS; Middleton et al., 2001)
Understanding problems and behaviours	Understanding behaviours	Items focused on therapists helping clients to gain an understanding of their problem behaviours, e.g., substance use.	<i>"To what extent did the therapist link the patient's drug use or abstinence to psychiatric symptoms?"</i> (CE Therapist Rating Form; Carroll et al., 1999) <i>"To what extent did the therapist help the patient to explore his/her feelings related to current symptoms or clarify affect states as related to drug use or other target problems?"</i> (YACSII; Nuro et al., 2005)
	Understanding relationships	Items referred to therapists helping clients to understand their interpersonal relationships, e.g., difficulties, similarities between past and present relationships.	<i>"Did the therapist focus on the client's interpersonal relationships?"</i> (UKATT PRS; Middleton et al., 2001) <i>"Did the therapist encourage the client to consider a broad range of potential options for dealing with an interpersonal problem (or role expectation issue within a relationship)?"</i> (CSPRS-6; SPR Project Staff, 1984)
	Understanding thoughts and feelings	Items focused on therapists increasing the clients' awareness of their thoughts and feelings, e.g., expression, exploration, and reality-testing.	<i>"Practitioner asks questions to elicit how patient thinks and feels about the topic."</i> (BECCI, Lane, 2002) <i>"Did the therapist explore with the client a general belief that underlies many of the client's specific negative thoughts and beliefs?"</i> (CSPRS-6; SPR Project Staff, 1984)
Resistance		Items related to therapists addressing the clients' resistance to change their behaviour and engage in treatment.	<i>"Rolling with resistance:</i> 1. <i>Argues with the client in favor of change</i> 4. <i>Notes client resistance without a change in own behavior/focus of session</i> 7. <i>Changes behavior/focus of session in order to reduce client resistance"</i> (MISTS Revised; Madson and Loignon, 2007)

Theme	Sub-theme	Description	Exemplar(s)
Wishes, dreams, and childhood		Items focused on therapists' exploring the clients' wishes, dreams or childhood experiences, and relating these issues to their current problems.	<p>"The therapist encourages discussion of patient's wishes, fantasies, dreams, or early childhood memories (positive or negative)." (CPPS; Hilsenroth et al., 2005)</p> <p>"To what extent did the therapist attempt to relate current problems to experiences occurring during the client's childhood or adolescence?" (MTRS; DeRubeis et al., 1982)</p>
Therapy specific techniques for increasing awareness		Items focused on therapy specific techniques, which help clients gain an understanding of their problems and behaviours, e.g., guided discovery.	<p>"_____ / _____ / _____ 22. Guided discovery (Adherence/Appropriateness/Quality)</p> <p><u>Quality rating:</u></p> <p>0 The therapist did not use guided discovery; instead was too passive or directive</p> <p>2 The therapist was somewhat passive or directive but was still supportive to the patient.</p> <p>4 The therapist used some questioning and some reflective responses to help pat. begin to understand important issues.</p> <p>6 The therapist very skilfully used a balance of open-ended questions, reflective, confrontative, and interpretive responses to guide patient's understanding of important issues." (CTACS; Liese et al., 1995)</p>

**BECCI** = Behaviour Change Counselling Index; **CBT Therapist Checklist** = Cognitive Behavioral Therapy Therapist Checklist; **CE Therapist Rating Form** = Compliance Enhancement Therapist Adherence/Competence Rating Form; **CPPS** = Comparative Psychotherapy Process Scale; **CSPRS-6** = Collaborative Study Psychotherapy Rating Scale – Form 6; **CTACS** = Cognitive Therapy Adherence and Competence Scale; **ITRS** = Independent Tape Rater Scale; **MISTS Revised** = Motivational Interviewing Supervision and Training Scale Revised; **MTRS** = Minnesota Therapy Rating Scale; **UKATT PRS** = UKATT (United Kingdom Alcohol Treatment Trial) Process Rating Scale; **YACSII** = Yale Adherence and Competence Scale Second Edition.



conceptualized as involuntarily operating mechanisms which the client uses as defences against anxiety, anger, shame, etc.?" (CSPRS-6; SPR Project Staff, 1984). Most were complex, asking multiple questions and/or including item-specific descriptive anchors (n=29); twenty-two of these items were coded from ACS-SEC:

*"The therapist engaged in an enactment of the central dynamic of the case.*

1. *The therapist clearly enacted the patient's central conflict without any sign of being aware of it.*
3. *The therapist enacted the patient's central conflict, but acknowledged part of it indirectly.*
5. *The therapist made the patient aware of the enactment of the conflict going on between them, and how it illuminates experiences with others.*
7. *As in 5, but with more subtlety and finesse; explored context and links to CCRT."*<sup>58</sup> (ACS-SEC; Barber, 1997)

#### **4.5.1.4 Interventions to change behaviour**

The fourth meta-theme, interventions to change behaviour, relates to the techniques therapists use to help the client change their behaviour and achieve their treatment goals (Table 12). Seven themes were included: i) advice, ii) encouraging behaviour change, iii) homework, iv) providing information, v) developing skills, vi) restructuring thoughts, and vii) therapy specific techniques.

##### **4.5.1.4.1 Advice**

Advice was the main focus for seven items, taken from seven measures; it refers to therapists' giving advice to the client. Three items focused on solicited advice: "*How often did the therapist give advice, make a suggestion, or offer a solution or possible action with prior permission from the client?*" (TPRS; Fisher et al., 2000). One item was concerned with unsolicited advice, therapists giving advice without the client's prior permission. The remaining three items were more general: "*Make suggestions regarding how or how much patient should cut down/stop?*" (BAS; Pantaloni et al., 2012)

##### **4.5.1.4.2 Encouraging behaviour change**

Behaviour change refers to the techniques therapists may use in which behaviour change is the salient feature. There were four sub-themes: i) talking about change, ii)

---

<sup>58</sup> CCRT = Core conflictual relationship theme.

commitment to change, iii) treatment goals, and iv) making a plan:

- i) **Talking about change:** This was the key feature for 13 items, coded from nine measures; it relates to therapists encouraging the client to talk about behaviour change. Five items considered change more generally: *“Practitioner invites the patient to talk about behaviour change”* (BECCI; Lane, 2002). Eliciting client optimism for change was addressed by two items: *“To what extent did the therapist attempt to elicit optimism for change from the client?”* (UKATT PRS; Middleton et al., 2001). The remaining six items referred to *“self-motivational statements”* or *“change talk”* – therapy specific terms, often used in motivational interviewing (MI), which refer to any discussion that favours change; this may include a conversation about the client’s desire, ability, reason, and need to change (Miller and Rollnick, 2013): *“The therapist guided the client toward verbalizing the need for change instead of telling the client why change was needed”* (GROMIT; Moyers, 2004)<sup>59</sup>.
- ii) **Commitment to change:** This was the main focus for three items, taken from three measures; it refers to therapists encouraging clients to make a commitment to change their drinking ( $n=2$ ), or drug use ( $n=1$ ): *“To what extent did the practitioner elicit a commitment from the client to change their drinking?”* (UKATT PRS; Middleton et al., 2001).
- iii) **Treatment goals:** This was the salient feature for 19 items, coded from 13 measures; it refers to therapists discussing the clients’ goals for treatment. Most items concentrated on therapists setting and reviewing treatment goals ( $n=8$ ): *“To what extent did the therapist review, set, or monitor patient goals for treatment?”* (CE Therapist Rating Form; Carroll et al., 1999). Eight items focused solely on goal setting, with five of these items emphasising the goal of abstinence: *“The interventionist advocates a goal of abstinence rather than reduction of drug use”* (IAC Treatment Fidelity Instrument; Torrey, 2011)<sup>60</sup>.
- iv) **Making a plan:** This was a relatively large sub-theme, encompassing 26

---

<sup>59</sup> GROMIT = Global Rating of Motivational Interviewing Therapist.

<sup>60</sup> IAC Treatment Fidelity Instrument = I Am Concerned Treatment Fidelity Instrument.

items, coded from 14 measures; it refers to therapists developing a plan with clients for making changes and achieving the clients' treatment goals. The sub-theme comprised several aspects, summarised by an extract from an ITRS item:

*"...To what extent did the clinician develop a change plan with the client in a collaborative fashion? How much did the clinician cover critical aspects of change planning such as facilitating a discussion of the client's self-identified goals, steps for achieving those goals, supportive people available to help the client, what obstacles to the change plan might exist..." (Martino et al., 2009b).*

Thirteen items focused on developing a plan: *"Did the therapist work with the client to schedule OR structure one or more specific activities for the purpose of increasing the likelihood that the client will initiate OR follow through on those activities"* (CSPRS-6; SPR Project Staff, 1984). Obstacles and potentials for achieving the clients' treatment goals was the main focus for one item (ACS-SEC; Barber et al., 1997).

A discussion of the clients' social network in relation to achieving their treatment goals was addressed by eight items; there was a distinction between identifying supportive people, and discussing the nature of that support: *"To what extent did you inquire about or discuss the availability and nature of family or social support for the patient's involvement in treatment or efforts to become abstinent?"* (CBT Therapist Checklist; Carroll, 1997). The remaining three items focused on relapse prevention and crisis planning. Two items, from ACS-IDCCD, were concerned with relapse: *"establishing concrete behavioral changes to get out of the relapse process"* (Mercer et al., 1995). Crisis planning was addressed in the CBT for PTSD Fidelity scale<sup>61</sup> (Lu et al., 2012).

#### 4.5.1.4.3 Homework

Homework was the salient feature for 39 items, coded from 14 measures; it focuses on therapists: i) assigning tasks for the client to do between sessions, and ii) reviewing previously assigned tasks. Nineteen items focused on task assignment. Most of these

---

<sup>61</sup> CBT for PTSD Fidelity Scale = Cognitive-Behavioral Therapy for Post-Traumatic Stress Disorder.

items were relatively general ( $n=10$ ): *“Did the therapist or client develop one or more specific assignment for the client to engage in between sessions?”* (SPRS; Shapiro and Startup, 1990). Other items specified a particular task ( $n=9$ ); for example, practising behaviours learned in therapy, attending 12-Step meetings, and monitoring thoughts associated with substance use:

*“Instruct the participant to keep a written “journal” to record 12-Step meetings attended (dates, time, locations) and personal reactions (thoughts, feelings, behaviors) about the meetings attended.”* (TSF-ACES; Campbell and Guydish, 2012)

Fifteen items focused on therapists reviewing previously assigned tasks. Similar to task assignment, most items did not specify the nature of the tasks reviewed ( $n=10$ ): *“To what extent did the therapist review concrete tasks assigned as homework during the previous therapy session?”* (ADAPTA PRS; Tober and Crosby, 2014)<sup>62</sup>. Other items referred to specific assignments ( $n=5$ ); for example, reviewing *“participant’s written journal from the last session”* (TSF-ACES; Campbell and Guydish, 2012). Seven items indicated how therapists might review previously assigned tasks: *“e.g., explore or address any difficulties encountered in carrying out the assignment, provide a rationale for homework, reinforce the importance of extra-session practice of skills”* (YACSII; Nuro et al., 2005).

The remaining five items combined task assignment and review. One of these items considered tasks in general: *“To what extent did the therapist plan or review concrete tasks assigned during therapy to be carried out outside therapy?”* (UKATT PRS; Middleton et al., 2001). Whereas four items focused on recording thoughts, feelings, activities or events:

*“Did the therapist encourage the client to record feelings, activities, or events between sessions OR review the client’s records of feelings, activities, or events?”* (CSPRS-6; SPR Project Staff, 1984).

#### 4.5.1.4.4 Providing information

Providing information was the main focus for 23 items, taken from 13 measures; it relates to therapists providing clients or family members with information. Three items did not

---

<sup>62</sup> ADAPTA PRS = ADAPTA (an alcohol-focused intervention versus a healthy living intervention for problem drinkers identified in a general hospital setting) Process Rating Scale.

specify the nature of the information: *“When practitioner provides information, it is sensitive to patient concerns and understanding”* (BECCI; Lane, 2002). Other items focused on a particular aspect of the clients’ substance use ( $n=16$ ); for example, withdrawal symptoms, relapse, prenatal substance use, and social support: *“The interventionist explains the effects that prenatal substance use can have on the mother, baby, and child.”* (IAC Treatment Fidelity Instrument; Torrey, 2011). The remaining three items highlighted related issues, including depression, high-risk behaviours, and trauma and post-traumatic stress disorder (PTSD):

*“Psychoeducation  
Information about trauma & PTSD  
Information about associated symptoms  
Elicit client’s symptoms  
Answer questions”* (CBT for PTSD Fidelity Scale; Lu et al., 2012)

The mode of delivery was addressed by 12 items: *“tell”, “explain”, “inform”, and “convey”* (e.g., YACSII; Nuro et al., 2005). Two of these items specified the use of educational material and information sheets: *“Give the patient the ED Information Sheet on health-risk behaviors”*<sup>63</sup> (BAS; Pantaloni et al., 2012).

#### 4.5.1.4.5 Developing skills

Developing skills was the key feature for 22 items, coded from 9 measures; it refers to therapists developing clients’ skills. Most items referred to skills more generally, providing examples to aid clarification ( $n=11$ ): *“Did the therapist attempt to teach the client skills (e.g., assertiveness, social skills, task relevant skills) in the session?”* (CSPRS-6; SPR Project Staff, 1984). Other items were more specific ( $n=11$ ): *“Helping the client to develop healthy social skills and/or recreational activities”* (ACS-IDCCD; Mercer et al., 1995). A range of techniques that therapists may use to facilitate skill development were specified ( $n=7$ ):

*“To what extent did you attempt to teach, model, rehearse, review or discuss specific skills (e.g., drug refusal, coping with craving, problem solving skills) during the session?”* (CBT Therapist Checklist; Carroll, 1997).

#### 4.5.1.4.6 Restructuring thoughts

This was the main focus for five items, taken from four measures; it relates to therapists’

---

<sup>63</sup> ED = Emergency department.

**Table 12: An overview of ‘interventions to change behaviour’ (meta-theme 4) with theme descriptions and exemplar items**

Theme	Sub-theme	Description	Exemplar(s)
Advice		Items related to therapists’ giving advice, or direction to clients.	<i>“How often did the therapist give advice, make a suggestion, or offer a solution or possible action with prior permission from the client?”</i> (TPRS; Fisher et al., 2000)
Encouraging behaviour change	Talking about change	Items focused on therapists encouraging clients to talk about change, including motivation to change.	<i>“Practitioner encourages patient to talk about change.”</i> (BECCI; Lane, 2002)  <i>“To what extent did the therapist attempt to elicit optimism for change from the client?”</i> (UKATT PRS; Middleton et al., 2001)
	Commitment to change	Items focused on therapists encouraging clients to make a commitment to change their substance use.	<i>“To what extent did the practitioner elicit a commitment from the client to change their drinking?”</i> (UKATT PRS; Middleton et al., 2001)
	Treatment goals	Items related to therapists discussing the clients’ goals for treatment, including setting and reviewing goals.	<i>“To what extent did the therapist discuss, review, or reformulate the patient’s goals for treatment?”</i> (YACSII; Nuro et al., 2005)
	Making a plan	Items focused on therapists developing a plan with clients for making changes and for achieving treatment goals.	<i>“To what extent did the therapist make a concrete behaviour change plan?”</i> (ADAPTA PRS; Tober and Crosby, 2014)
Homework		Items were concerned with therapists assigning tasks for clients to do between sessions, or reviewing previously assigned tasks.	<i>“Did you develop one or more specific assignments for the patient to engage in between sessions?”</i> (CBT Therapist Checklist; Carroll, 1997)  <i>“To what extent did the therapist review the patient’s reactions to last session’s assignment (e.g., explore or address any difficulties encountered in carrying out the assignment, provide a rationale for homework, reinforce the importance of extra-session practice of skills)?”</i> (YACSII; Nuro et al., 2005)

Theme	Sub-theme	Description	Exemplar(s)
Providing information		Items related to therapists providing information to clients or family members.	<i>"The therapist provides the patient with information and facts about his or her current symptoms, disorder, or treatment."</i> (CPPS; Hilsenroth et al., 2005)
Developing skills		Items were concerned with therapists developing the clients' skills. A range of therapist techniques were specified, e.g., teaching, practicing, and reviewing skills.	<i>"Did the therapist attempt to teach the client skills (e.g., assertiveness, social skills, task relevant skills) in the session?"</i> (CSPRS-6; SPR Project Staff, 1984)
Restructuring thoughts		Items focused on changing clients' thoughts and beliefs.	<i>"Did the therapist and client practice possible rational responses to the client's negative thoughts or beliefs?"</i> (CSPRS-6; SPR Project Staff, 1984)
Therapy specific techniques for changing behaviour		Items related to therapy specific techniques for changing clients' behaviour.	<i>"Did the CM clinician state how many draws would be earned at the next session if patient were abstinent?"</i> (CM Clinician Rating Form; Petry and Stitzer, 2002)

**ADAPTA PRS** = ADAPTA (an alcohol-focused intervention versus a healthy living intervention for problem drinkers identified in a general hospital setting) Process Rating Scale; **BECCI** = Behaviour Change Counselling Index; **CBT Therapist Checklist** = Cognitive Behavioral Therapy Therapist Checklist; **CM Clinician Rating Form** = Contingency Management Clinician Adherence/Competence Rating Form; **CPPS** = Comparative Psychotherapy Process Scale; **CSPRS-6** = Collaborative Study Psychotherapy Rating Scale – Form 6; **TPRS** = Therapy Process Rating Scale; **UKATT PRS** = UK Alcohol Treatment Trial Process Rating Scale; **YACSII** = Yale Adherence and Competence Scale Second Edition.

attempts to change clients' thoughts and beliefs. A range of therapist techniques were addressed by one item:

*“Cognitive Restructuring  
Thought-feeling model  
Connect negative feelings to thoughts  
Challenge thoughts  
Generate alternative thoughts  
Practice alternative thoughts” (CBT for PTSD Fidelity Scale; Lu et al., 2012)*

The remaining four items focused on one particular technique: *“Did the therapist and client practice possible rational responses to the client's negative thoughts or beliefs?”* (CSPRS-6; SPR Project Staff, 1984).

#### *4.5.1.4.7 Therapy specific techniques for changing behaviour*

This was the salient feature for nine items, taken from seven measures; it refers to specific therapeutic techniques for changing clients' behaviour. There were items specific to CBT ( $n=1$ ), contingency management (CM;  $n=4$ ), and MI ( $n=4$ ): *“Did the CM clinician state how many draws were earned at this session?”* (CM Clinician Rating Form; Petry and Stitzer, 2002).

#### **4.5.1.5 Core skills**

The fifth meta-theme, core skills, relates to how therapists delivered the session (**Table 13**). There were nine themes: i) supporting client self-efficacy, ii) collaboration, iii) empathic relationship, iv) presentation, v) self-disclosure, vi) asking questions, vii) reflective listening, viii) overall performance, and ix) negative therapist attributes.

##### *4.5.1.5.1 Supporting client self-efficacy*

Supporting client self-efficacy was the main focus for 19 items, taken from 12 measures; it relates to therapists recognising and reinforcing the clients' strengths, abilities, or efforts to change. Three items focused on the clients' strengths: *“The therapist directed the client's attention toward their own strengths”* (GROMIT; Moyers, 2004). Other items were concerned with praising clients, and encouraging their efforts to change ( $n=6$ ): *“To what extent did the therapist compliment and/or praise a past patient effort that did not include the role of medication?”* (YACSII; Nuro et al., 2005). In cases where change had not yet occurred, items focused on therapists communicating a sense of optimism about client



change ( $n=6$ ): *“To what extent did the CM clinician communicate confidence that patient's efforts will yield success in the future?”* (CM Clinician Rating Form; Petry and Stitzer, 2002). The remaining four items were multifaceted:

*“To what extent did the therapist verbally reinforce the client's strengths, abilities, or efforts to change his/her behavior? To what extent did the therapist encourage a sense of self-efficacy on the part of the client by praising small steps in the direction of change or expressing appreciation of personal qualities in the client that might facilitate successful efforts to change?”* (YACSII; Nuro et al., 2005)

#### 4.5.1.5.2 Collaboration

Collaboration was the salient feature for 29 items, coded from 15 measures; it refers to the therapist and the client working together. Five items were more general: *“Did the therapist actively attempt to engage the client in working together to explore therapeutic issues?”* (CSPRS-6; SPR Project Staff, 1984). Other items indicated how therapists might foster a collaborative relationship ( $n=11$ ). There were a variety of ways to encourage collaboration; for example, *“using the language of shared endeavour (‘I’ and ‘we’)”* (SPRS; Shapiro and Startup, 1990), engendering *“curiosity”* (ACS-SEC; Barber, 1997), and avoiding *“an expert/authoritarian role”* (MISTS Revised; Madson and Laignon, 2007). YACSII and ITRS included a similar item, providing a more comprehensive view of collaboration:

*“To what extent did the therapist convey in words or actions that the therapy is a collaborative relationship in contrast to one where the therapist is in charge? How much did the therapist emphasize the (greater) importance of the client's own decisions, confidence, and perception of the importance of changing? To what extent did the therapist verbalize respect for the client's autonomy and personal choice?”* (YACSII; Nuro et al., 2005)

The remaining 13 items focused on a particular aspect of collaborative working; for example, encouraging client independence ( $n=4$ ) and negotiating the content of therapy ( $n=3$ ): *“Did the therapist encourage the client's independence from the therapist in dealing with her/his problems?”* (CSPRS-6; SPR Project Staff, 1984).

#### 4.5.1.5.3 Empathic relationship

Empathic relationship relates to therapists communicating understanding of and sensitivity to the client's perspective. There were seven sub-themes: i) empathy, ii) acceptance, respect, and being non-judgemental, iii) support and reassurance, iv) warmth and

genuineness, v) being attuned and attentive, vi) rapport, and vii) sensitivity and concern.

- i) **Empathy:** This was the main focus for 14 items, taken from 13 measures; it concentrates on the extent to which therapists respond empathetically to the clients: *“Practitioner uses empathic listening statements when the patient talks about the topic”* (BECCI; Lane, 2002). Eleven of these items indicated how therapists might convey empathy: *“...through a non-judgemental stance, showing genuine warmth and concern, helping the patient feel accepted in the relationship”* (CBT Therapist Checklist; Carroll, 1997). Items coded as empathy included elements from the other sub-themes.
  
- ii) **Acceptance, respect, and being non-judgemental:** This was a key feature for 13 items, coded from eight measures; it refers to therapists conveying respect, appearing non-judgemental and accepting of the clients’ experiences. Most items focused on respect ( $n=6$ ): *“Practitioner actively conveys respect for patient choice about behaviour change”* (BECCI; Lane, 2002). Appearing non-judgemental and accepting was addressed by three items: *“How much is the clinician accepting of the client’s feelings and inner experiences?”* (ITRS; Martino et al., 2009b). The remaining four items were multifaceted; for example, a high quality rating in CTACS would be achieved if *“the therapist appeared fully accepting, respectful and nonjudgemental”* (Liese et al., 1995).
  
- iii) **Warmth and genuineness:** This was the main focus for six items, taken from five measures; it refers to therapists conveying warmth and genuineness. Most items focused on the former ( $n=4$ ): *“How much did the clinician convey warmth?”* (ITRS; Martino et al., 2009b). One item looked at whether *“the therapist appeared to be 1-sincere and genuine [or] 9-insincere or contrived”* (MTRS; DeRubeis et al., 1982). The remaining item was multifaceted; to receive a high quality rating in CTACS, therapists must appear *“optimally warm, genuine, caring, and congruent”* (Liese et al., 1995).
  
- iv) **Support and reassurance:** This was the key feature for eight items, coded from four measures; it focuses on therapists being supportive and reassuring to the clients. Most items focused on therapists providing support ( $n=6$ ); five of these items were coded from the IAC Treatment Fidelity Scale: *“The interventionist*

*uses a supportive tone of voice when responding to the woman's reaction"* (Torrey, 2011). The extent to which therapists provided reassurance was addressed by one item (CE Therapist Rating Form; Carroll et al., 1999). The remaining item combined both features: *"Rate the extent to which the therapist attempted to be supportive or reassuring"* (TPRS; Fisher et al., 2000).

- v) ***Being attuned and attentive***: This was the main aspect for eight items, taken from seven measures; it concentrates on therapists being attuned and attentive to the clients. Most items were concerned with attentiveness ( $n=4$ ); for example, a high quality rating in CTACS was achieved if: *"The therapist was extremely attentive to important obvious and subtle cues"* (Liese et al., 1995). The remaining three items looked at whether therapists were attuned to the clients' feelings ( $n=2$ ), or inner world ( $n=1$ ): *"How much is the clinician attuned to the client's inner world moment by moment in the session?"* (ITRS; Martino et al., 2009b).
- vi) ***Rapport***: This was the main focus for four items, taken from four measures; it relates to *"How much rapport was there between therapist and client (i.e., how well did the therapist and client get along)?"* (CSPRS-6; SPR Project Staff, 1984).
- vii) ***Sensitivity and concern***: This was a key feature for five items, coded from five measures; it refers to therapists expressing concern, and sensitivity to the clients' emotions, negative reactions, and other issues: *"Practitioner demonstrates sensitivity to talking about other issues"* (BECCI; Lane, 2002).

#### 4.5.1.5.4 Presentation

Presentation was the main focus for 16 items, taken from six measures; it refers to therapists' style of presentation. Most items focused on therapists level of verbal and non-verbal communication ( $n=10$ ). For example, three items were concerned with the level of verbal activity during the session: *"Who talked the most during the session? 1-therapist 5-equivalent 9-client"* (MTRS; DeRubeis et al., 1982). Other items considered therapists' level of expressiveness, verbally ( $n=4$ ) and non-verbally ( $n=1$ ): *"How interesting is the therapist's style of communication? (Consider (1) the vividness of her/his language; (2) the originality of her/his ideas; (3) the liveliness of her/his manner of speaking)"* (CSPRS-6; SPR Project Staff, 1984). The remaining six items related to therapists level of professionalism and expertise.

Items considered whether therapists' exuded confidence ( $n=1$ ), competence ( $n=3$ ), and formality<sup>64</sup> ( $n=2$ ): *"The therapist presents him/herself in a professional and competent manner"* (TPRS; Fisher et al., 2000).

#### 4.5.1.5.5 Self-disclosure

Self-disclosure was the salient feature for two items, coded from two measures; it focuses on therapists' use of appropriate self-disclosure, including therapists' own experiences and beliefs: *"Did the therapist make appropriate use of self-disclosure to advance a shared understanding of events and processes in the therapy?"* (SPRS; Shapiro and Startup, 1990). Items referring to inappropriate disclosure were coded as negative therapist attributes.

#### 4.5.1.5.6 Asking questions

Asking questions was the main focus for 11 items, coded from eight measures; it refers to therapists asking questions. Most items looked at the use of open-ended questions ( $n=7$ ): *"To what extent did the therapist use open-ended questions (i.e., questions that elicit more than yes/no responses) to elicit the client's perception of his/her problems, motivation, change efforts, and plans?"* (YACSII, Nuro et al., 2005). The use of closed-ended questions was addressed by 3 items: *"To what extent did the clinician ask questions that could be answered with a yes or no response or that sought after specific details or information from the client?"* (ITRS; Martino et al., 2009b). The remaining item referred to both open-ended and closed-ended questions: a high quality rating would be given in MISTS Revised if there was *"Good facilitation of client exploration through the use of primarily open questions"* (Madson and Loignon, 2007).

#### 4.5.1.5.7 Reflective listening

Reflective listening was the salient feature for 24 items, taken from 20 measures; it relates to therapists' attempts to communicate understanding of what clients have said. Most items were relatively general ( $n=15$ ): *"Overall, how well did the counselor understand or make an effort to grasp the client's perspective?"* (TSF-ACES; Campbell and Guydish, 2012). Five of these items indicated how therapists may use reflective statements; for example, repeating the client's comments (simple reflection), and adding new meaning (complex reflection):

---

<sup>64</sup> Formality is defined as *"strict adherence to the therapeutic role such that little of the therapist's own personality emerges during the session."* (SPR Project Staff, 1984, p.36)

*“To what extent did the clinician use simple reflections by repeating (exact words) or rephrasing (slight rewording) what the client had said? To what extent did the clinician use complex reflections by paraphrasing what the client had said (i.e., stating what was implied, thereby adding related, yet new meaning to the client’s statements)?” (ITRS; Martino et al., 2009b)*

The remaining items were more specific, focusing on the client’s feelings ( $n=2$ ), emotions ( $n=2$ ), and thoughts ( $n=5$ ): *“How often did the therapist provide reflections of the client’s emotional states during the session?”* (TPRS; Fisher et al., 2000).

#### *4.5.1.5.8 Overall performance*

Overall performance was the main focus for 24 items, coded from 14 measures; it refers to the assessment of therapists’ overall performance, considering the appropriateness of the therapeutic techniques delivered, the level of skill shown by the therapist, and the quality of the session. Most items were concerned with therapists’ use of appropriate techniques during the session ( $n=9$ ): *“Appropriate use of Integrated MI and CBT strategies and techniques”* (MI-CTS; Haddock et al., 2012). Eight items focused on therapists’ general level of skilfulness and/or effectiveness. The remaining seven items assessed the overall session quality; for example, a high rating in CTACS would be achieved if:

*“The therapist’s performance in this session is excellent. Cognitive therapy is practiced at a level equal to or superior to supervisor’s own level of proficiency. Therapist apparently knows the treatment manual extremely well. Applies the cognitive case formulation with ease and flexibility. This represents “state of the art” cognitive therapy.”* (Liese et al., 1995)

#### *4.5.1.5.9 Negative therapist attributes*

Negative therapist attributes encapsulated 20 items, taken from 9 measures; it refers to therapist attributes considered detrimental to therapeutic progress and behaviour change. Most items focused on therapist confrontation ( $n=7$ ): *“To what extent did the therapist explicitly confront the patient and/or employ a confrontational style working with the patient?”* (YACSII; Nuro et al., 2005). Other negative attributes included: expressing frustration and disapproval ( $n=2$ ), appearing inattentive and disingenuous ( $n=3$ ), and being argumentative and overly directive ( $n=6$ ): *“How often did the therapist give an order, command, or direction?”* (CPPS; Hilsenroth et al., 2005). The remaining two items were concerned with therapists’ use of inappropriate techniques, and engagement in unrelated discussions and self-disclosure:

**Table 13: An overview of ‘core skills’ (meta-theme 5) with theme descriptions and exemplar items**

Theme	Sub-theme	Description	Exemplar(s)
Supporting client self-efficacy		Items related to therapists recognising and reinforcing the client’s strengths, abilities, or efforts to change.	<i>“To what extent did the therapist verbally reinforce the client’s strengths, abilities, or efforts to change his/her behavior? To what extent did the therapist encourage a sense of self-efficacy on the part of the client by praising small steps in the direction of change or expressing appreciation of personal qualities in the client that might facilitate successful efforts to change?”</i> (YACSII; Nuro et al., 2005)
Collaboration		Items were concerned with the therapist and the client working together.	<i>“Did the therapist actively attempt to engage the client in working together to explore therapeutic issues?”</i> (CSPRS-6; SPR Project Staff, 1984)
Empathic relationship	Empathy	Items related to the extent to which therapists conveyed empathy.	<i>“To what degree did you respond empathically to the patient (e.g., through a non-judgemental stance, showing genuine warmth and concern, helping the patient feel accepted in the relationship)?”</i> (CBT Therapist Checklist; Carroll, 1997)
	Acceptance, respect, and being non-judgemental	Items were concerned with therapists conveying respect, and appearing non-judgemental and accepting of the client.	<p><i>“ _____ / _____ / _____ 12. Acceptance/respect (Adherence/Appropriateness/Quality)</i></p> <p><i>Quality rating:</i></p> <p><i>0 The therapist appeared critical, disrespectful, judgmental.</i></p> <p><i>2 The therapist appeared slightly critical, disrespectful, judgmental.</i></p> <p><i>4 The therapist appeared reasonably accepting, respectful, nonjudgmental.</i></p> <p><i>6 The therapist appeared fully accepting, respectful, nonjudgmental.”</i></p> <p>(CTACS; Liese et al., 1995)</p>
	Warmth and genuineness	Items focused on therapists conveying warmth and genuineness.	<i>“Did the therapist convey warmth?”</i> (CSPRS-6; SPR Project Staff, 1984)

Theme	Sub-theme	Description	Exemplar(s)
Empathic relationship continued	Support and reassurance	Items related to therapists being supportive and reassuring.	<i>"Rate the extent to which the therapist attempted to be supportive or reassuring."</i> (TPRS; Fisher et al., 2000)
	Being attuned and attentive	Items focused on therapists being attuned and attentive to what the client was saying.	<i>"The therapist is able to attune to the patient's feelings."</i> (TPRS; Fisher et al., 2000)
	Rapport	Items were concerned with how much rapport there was between the therapist and the client.	<i>"How much rapport was there between therapist and client? 1-excellent rapport 9-absence of rapport"</i> (MTRS; DeRubeis et al., 1982)
	Sensitivity and concern	Items focused on therapists expressing concern, and being sensitive to the client's emotions, negative reactions, and issues.	<i>"How much did the clinician demonstrate concern for the client?"</i> (ITRS; Martino et al., 2009b) <i>"Practitioner demonstrates sensitivity to talking about other issues"</i> (BECCI; Lane, 2002)
Presentation		Items referred to therapists' style of presentation, including: i) therapists' verbal and non-verbal activity, and ii) level of professionalism and expertise.	<i>"How much did the therapist talk?"</i> (CSPRS-6; SPR Project Staff, 1984) <i>"How much did the clinician demonstrate expressiveness in his or her voice?"</i> (ITRS; Martino et al., 2009b) <i>"The therapist presents him/herself in a professional and competent manner."</i> (TPRS; Fisher et al., 2000)
Self-disclosure		Items related to therapists' use of self-disclosure, e.g., telling the client about their own experiences and beliefs.	<i>"To what extent did the therapist explicitly refer to his/her own life experiences or beliefs?"</i> (YACSII; Nuro et al., 2005)

Theme	Subtheme	Description	Exemplar(s)
Asking questions		Items were concerned with therapists asking the client open questions (eliciting more than a yes/no response) and closed questions (eliciting a yes/no response).	<p><i>“To what extent did the therapist use open-ended questions (i.e., questions that elicit more than yes/no responses) to elicit the client’s perception of his/her problems, motivation, change efforts, and plans?”</i> (YACSII; Nuro et al., 2005)</p> <p><i>“How often did the therapist ask closed-ended questions during the interview?”</i> (TPRS; Fisher et al., 2000)</p>
Reflective listening		Items related to therapists’ attempts to communicate understanding of what the client said, e.g., repeating client’s comments (simple reflections), and adding new meaning to what the client said (complex reflections).	<p><i>“To what extent did the clinician use simple reflections by repeating (exact words) or rephrasing (slight rewording) what the client had said? To what extent did the clinician use complex reflections by paraphrasing what the client had said (i.e., stating what was implied, thereby adding related, yet new meaning to the client’s statements)?”</i> (ITRS; Martino et al., 2009b)</p>
Overall performance		Items focused on the therapists overall performance by considering the appropriateness of the therapeutic techniques delivered, the level of skill shown by the therapist, and the quality of the session.	<p><i>“Overall, how well did the counselor conduct this specific individual session?”</i> (TSF-ACES; Campbell and Guydish, 2012)</p>
Negative therapist attributes		Items referred to therapist attributes considered detrimental to therapeutic progress and behaviour change, e.g., expressing frustration.	<p><i>“To what extent did the therapist explicitly confront the patient and/or employ a confrontational style working with the patient?”</i> (YACSII; Nuro et al., 2005)</p>

**BECCI** = Behaviour Change Counselling Index; **CBT Therapist Checklist** = Cognitive Behavioral Therapy Therapist Checklist; **CSPRS-6** = Collaborative Study Psychotherapy Rating Scale – Form 6; **CTACS** = Cognitive Therapy Adherence and Competence Scale; **ITRS** = Independent Tape Rater Scale; **MTRS** = Minnesota Therapy Rating Scale; **TPRS** = Therapy Process Rating Scale; **TSF-ACES** = Twelve Step Facilitation Adherence Competence Empathy Scale; **YACSII** = Yale Adherence and Competence Scale Second Edition.



*“To what extent did the clinician speak with the client about topics that were not related to the problems for which the client entered treatment? To what extent did the clinician disclose personal information about him/herself that was not solicited by the client or used to clarify the client’s own experiences?” (ITRS; Martino et al., 2009b)*

#### **4.5.2 Stage 2: Selecting themes for the BATS**

So far, this chapter has focused on developing an understanding of the range of therapist activities described by the items from the identified measures. Items were grouped based on what aspect of therapeutic practice they targeted. Thirty three themes were developed and were grouped into five meta-themes. The next part of this chapter focuses on the themes or activities most relevant to the BATS.

In order to select themes most relevant to the BATS, the thematic structure (developed in stage 1) was refined in collaboration with the supervision team; the number of themes was reduced from 33 to 18 (**Table 14**). The rationale for retaining, revising, and removing themes is provided in Appendix C; an overview of the changes is given below.

- i) **Session management** (meta-theme 1): ‘Focus and structure’ sub-themes were combined to form a new theme ‘problem focused’. The remaining themes were removed; the content was considered specific to a therapeutic approach/session, or indistinct from the other themes. For example, ‘session content’ was removed, the coded items focused on the particular therapeutic approach or the specific techniques delivered during the session.
- ii) **Medication and case management** (meta-theme 2): All themes were removed; that is, the themes were not selected, as they were not considered relevant to the BATS. ‘Medication’ was not considered transtheoretical, and ‘case management’ was not sufficiently distinct from the other themes. For example, as part of case management, a therapist and client may plan for the client to attend a self-help group between therapy sessions. This example would be captured by the homework themes.
- iii) **Interventions to increase awareness** (meta-theme 3): Sub-themes for ‘explores behaviours’ were retained with minor name changes, i.e., ‘exploring behaviours’ and ‘exploring substance use’. ‘Ambivalence’ was split into two

**Table 14: Summary of the revised thematic structure**

Meta-theme	Theme	Sub-themes	Refined Theme	Description of the Refined Theme
Session management	Focus and structure	Maintaining structure Agenda setting Consistency of problem focus Directiveness Time management	Problem focused	Therapists keeping the session focused on the aims for that session.
Interventions to increase awareness	Explores behaviours	Explores general behaviours	Exploring behaviours	Therapists exploring the clients' problem behaviours without reference to substance use or other concerns.
		Explores substance use	Exploring substance use	Therapists' exploring the clients' substance use problems.
	Ambivalence	Exploring Impact of substance use	Therapists' exploring the impact of the client's substance use.	
		Exploring pros and cons of change	Therapists increasing the clients' awareness of what might be gained and lost by changing their substance use.	
	Creates conflict	Developing discrepancy	Therapists heightening discrepancies experienced by clients' on conflicting aspects of their life.	
Interventions to change behaviour	Encouraging behaviour change	Talking about change	Talking about change	Therapists encouraging clients to talk about change.
		Treatment goals	Treatment goals	Therapists discussing the clients' goals for treatment, including setting and reviewing agreed goals.

Meta-theme	Theme	Sub-theme	Revision(s)	Description of revised themes
Interventions to change behaviour	Behaviour change	Making a plan	Behaviour change planning	Therapists developing and reviewing a plan with clients for making changes and achieving the clients' treatment goals.
			Identifying sources of support	Therapists identifying supportive people available to help the clients in making changes and achieving the clients' treatment goals.
			Involving others	Therapists discussing how the clients' social network may support them in making changes and achieving the clients' treatment goals.
	Homework		Home assigned	Therapists assigning tasks for clients' to do between sessions.
			Homework reviewed	Therapists reviewing previously assigned tasks.
Core skills	Supporting client self-efficacy		Strengths and affirmations	Therapists recognising and reinforcing the clients' strengths, abilities, or efforts to change.
	Collaboration		Collaboration	Therapists working together with the clients.
	Empathic relationship	Empathy	Empathy	Therapists conveying warmth and understanding of the clients' thoughts and feelings. Therapists appear non-judgemental and supportive of the client.
	Reflective listening		Simple reflections	Therapists repeating or slightly rephrasing what clients have said.
			Complex reflections	Therapists paraphrasing the clients' comments to add meaning.

themes: 'exploring impact of substance use' and 'exploring pros and cons of change'. The theme 'ambivalence' was revised because the coded items evaluated two distinct therapist behaviours. 'Creates conflict' was retained with a change of name; such item behaviours usually succeed therapists' exploration of the pros and cons of change. The name was changed to 'developing discrepancy' to better reflect the theme definition. Themes that did not fit with this refined meta-theme were removed; the content was considered therapy specific, or was captured by other themes. For example, there were no obvious exemplars for the 'understanding relationships' sub-theme, a range of behaviours were evaluated. Thus, items were captured by 'exploring impact of substance use' and the 'reflective listening' themes.

- iv) ***Interventions to change behaviour*** (meta-theme 4): All four sub-themes for 'encouraging behaviour change' were revised. First, 'moving towards change' was retained with no changes. Second, 'commitment to change' was removed (not selected); the content was captured by the other 'behaviour change' sub-themes. Third, 'treatment goals' was retained with no changes. Lastly, 'making a plan' was split into three groups as multiple therapist behaviours were covered: 'behaviour change planning', 'identifying sources of support', and 'involving others'. The 'homework' theme was also split into two themes: 'homework assigned', and 'homework reviewed'. The remaining themes were removed; the content was considered therapy specific, or was captured by other themes.
- v) ***Core skills*** (meta-theme 5): There were three revisions. First, 'supporting self-efficacy' was retained with a change in name, i.e., 'strengths and affirmations'. Second, only one sub-theme for 'empathic relationship' was retained with no changes; the six remaining sub-themes were removed, the coded items were captured by the retained sub-theme 'empathy'. Third, 'reflective listening' was split into two themes: 'simple reflections' and 'complex reflections'. The theme 'collaboration' was retained with no changes. The remaining themes were removed; the content was captured by other themes. For example, 'self-disclosure' concerned therapists' use of appropriate self-disclosure, a technique therapists may use to convey 'empathy' and foster 'collaboration'. The theme 'negative therapist attributes' was removed as the BATS was to focus on prescribed behaviours.

### 4.5.3 Stage 3: Generating items for the BATS

Items for the scale were based on the refined thematic structure, i.e., selected themes considered by the researcher and the supervision team to be most relevant to the BATS. Eighteen items were developed (Table 15). Theme names were used as the item reference. Most items were adapted from coded items taken from the identified measures, eight items were newly constructed.

### 4.5.4 Stage 4: Generating response formats for the BATS

Two response formats for potential inclusion in the BATS were generated: the first measured extensiveness, and the second measured quality (Figure 5). The formats reflected existing methods for assessing treatment adherence and therapist competence. Most of the measures identified in Study 1 asked raters to score: i) the extent to which therapists carried out item specific behaviours (treatment adherence), and ii) the quality with which therapists performed the behaviours (therapist competence).

Extensiveness (*"To what extent..."*)

*"To what extent did the therapist cover both the pros and cons for change?"*

0	1	2	3	4
Not at all	A little	Somewhat	Considerably	Extensively

Quality (*"How well..."*)

*"How well did the therapist cover both the pros and cons for change?"*

0	1	2	3	4
Very poor	Poor	Good enough	Well	Very well

**Figure 5: Response formats generated for the BATS**

The formats both used 5-point scales. An odd number of response options was consistent with the scoring methods used by the identified measures. While 7-point scales were most common ( $n=15$ ), only four measures provided descriptive anchors for all of the points. Streiner et al. (2015) recommend that all response options should be labelled because raters tend to score labelled anchors more frequently than unlabeled ones, and user satisfaction is improved. Less than five response options was not considered appropriate.

**Table 15: Generated items for potential inclusion in the BATS**

Meta-theme	Item reference	Item	Item type
Session management	1. Problem focused	To what extent did the therapist organise the session so that defined tasks were covered?	New
Interventions to increase awareness	2. Exploring behaviours	To what extent did the therapist encourage the client to talk about their current behaviour or status quo?	Adapted from BECCI (Lane, 2002)
	3. Exploring substance use	To what extent did the therapist gather information on the client's past or recent use of alcohol and drugs (e.g., patterns of use, extent of urges/thought, extent of reduction in use, results of recent urine/breath tests)?	Adapted from ITRS (Martino et al., 2009b)
	4. Exploring Impact of substance use	To what extent did the therapist explore the positive and negative aspects of the client's substance use?	New
	5. Exploring pros and cons of change	To what extent did the therapist cover both the pros and cons for change?	New
	6. Developing discrepancy	To what extent did the therapist encourage the client to contrast their addictive behaviour with personal goals or values?	New
Interventions to change behaviour	7. Talking about change	To what extent did the therapist encourage the client to talk about the reasons to change their substance use behaviour?	New
	8. Treatment goals	To what extent did the therapist develop or review the client's goals for treatment?	New
	9. Behaviour change planning	To what extent did the therapist develop or review a change plan (e.g., steps, possible obstacles, sources of support, and solutions)?	New

Meta-theme	Item reference	Item	Item type
Interventions to change behaviour continued	10. Identifying sources of support	To what extent did the therapist discuss the client's social network, exploring the nature of the client's relationships with network members and how those relationships might be used to support recovery?	Adapted from YACSII (Nuro et al., 2005)
	11. Involving others	To what extent did the therapist discuss the availability of specific individuals who will be or are sources of support for the client's involvement in treatment or efforts to change their substance use behaviour?	Adapted from UKATT PRS (Middleton et al., 2001)
	12. Home assigned	To what extent did the therapist with the client to plan specific tasks for the client to engage in between sessions?	Adapted from CSPRS-6 (SPR Project Staff, 1984)
	13. Homework reviewed	To what extent did the therapist review previously assigned homework with the client?	Adapted from CSPRS-6 (SPR Project Staff, 1984)
Core skills	14. Strengths and affirmations	To what extent did the therapist focus on the client's strengths, abilities, or efforts to change?	Adapted from YACSII (Nuro et al., 2005)
	15. Collaboration	To what extent did the therapist actively attempt to engage the client in working together to explore therapeutic issues?	Adapted from CSPRS-6 (SPR Project Staff, 1984)
	16. Empathy	To what extent did the therapist communicate understanding of and sensitivity to the client's comments and concerns?	New
	17. Simple reflections	To what extent did the therapist use simple reflections by repeating (exact words or with slight rewording) what the client had said?	Adapted from ITRS (Martino et al., 2009b)
	18. Complex reflections	To what extent did the therapist use complex reflections by paraphrasing what the client had said (e.g., adding new meaning and enabling the client to make connections)?	Adapted from ITRS (Martino et al., 2009b)

**BECCI** = Behaviour Change Counselling Index; **CSPRS-6** = Collaborative Study Psychotherapy Rating Scale – Form 6; **ITRS** = Independent Tape Rater Scale; **UKATT PRS** = UKATT (United Kingdom Alcohol Treatment Trial) Process Rating Scale; **YACSII** = Yale Adherence and Competence Scale Second Edition.

Fewer options can impact on the raters' ability to "*discriminate differences in the underlying attribute*", thereby reducing reliability of the scale (Streiner et al., 2015; DeVellis, 2017, p.123). Item-specific anchors were not included, the BATS was designed to be brief and easy to use. Keeping the labels the same for each item reduced the burden on the scale users (Streiner et al., 2015). A 'not applicable' option was not included to encourage raters to make a judgement on each item.

## **4.6 Discussion**

### **4.6.1 Study overview**

Measures identified from the literature review in Study 1 were used as a basis for generating items for potential inclusion in the BATS, and for informing a decision about how the items should be scored. Items from the identified measures were analysed using a form of thematic analysis (Braun and Clarke, 2006); items were grouped based on what aspect of therapeutic practice they targeted. The thematic structure was refined, in collaboration with the supervision team, to identify the themes most relevant to the BATS. Themes concerning therapy specific techniques were not prioritised, as the BATS was to be transtheoretical. The refined themes were expressed as individual items for potential inclusion in the BATS. The items were chosen to reflect the key features of therapies widely used in the treatment of alcohol and drug use. Consideration was also given at this stage to scoring. Possible response formats for the BATS were generated, these were informed by the identified measures.

### **4.6.2 Main findings**

The main findings for this chapter focus on the items and the response formats generated for the BATS.

#### **4.6.2.1 Items for the BATS**

The results of the thematic analysis identified 33 themes, which were grouped into five-meta themes. The first meta-theme, session management, focused on the techniques therapists may use to manage the therapy session. The second meta-theme, medication and case management, focused on therapists' discussion of the clients' medication, and involvement in self-help groups and other services. The third meta-theme, interventions to increase awareness, concentrated on the techniques therapists may use to increase clients' understanding of their behaviours, thoughts, feelings and relationships. The fourth



meta-theme, interventions to change behaviour, related to the techniques therapists may use to help clients change their behaviour and achieve their treatment goals. The last meta-theme, core skills, related to therapists' style, how therapists delivered the session. The refined thematic structure identified 18 themes most relevant to the BATS. Items were chosen from each of the themes. Most of the items were adapted from the extracted items on the identified measures; eight were newly constructed but reflected the aspect of practice targeted by the original measures. The 18 exemplar items reflected the key features of therapies widely used in the treatment of alcohol and drug use.

#### **4.6.2.2 Response formats for the BATS**

Two response formats were generated: the first measured extensiveness, and the second measured quality. The formats reflected existing methods for assessing treatment adherence and therapist competence. Most of the measures identified in Study 1 asked raters to score: i) the extent to which therapists carried out item specific behaviours (treatment adherence), and ii) the quality with which therapists performed the behaviours (therapist competence).

#### **4.6.3 Strengths and limitations**

A key strength of this study was the comprehensive and transparent account of the analytic process (Mays and Pope, 2006). Thematic analysis was used to group items based on what aspect of therapeutic practice they were targeting. The analysis provides "*a rich and detailed, yet complex account*" of the data (Braun and Clarke, 2006, p.78); key features of therapies widely used in the addiction field are highlighted. The thematic structure was developed in collaboration with the supervision team, reducing the potential bias that comes from a single researcher (Patton, 2002); alternative explanations and organising schemes were considered, and the credibility of the findings was enhanced (Patton, 2002; Nowell et al., 2017). Rigour could have been improved by having two (or more) researchers independently analyse the data and compare the findings (Patton, 2002; Joffe, 2012). Analysis triangulation is useful when conducting complex thematic analysis, as "*coding decisions are made explicit and consistent*" (Joffe and Yardley, 2004, p.63). However, items from the identified measures were embedded within the analytic narrative. Including example items demonstrates the rigour of the analysis (Braun and Clarke, 2006); the interpretation offered is supported by the data (Mays and Pope, 2006; Willig, 2013).

The thematic structure (developed in stage 1) was refined, in collaboration with the supervision team, to identify themes most relevant to the BATS. The supervision team was highly experienced in the addiction and psychotherapy fields, lending credibility to the findings. The diversity of backgrounds was advantageous; group heterogeneity can enhance the decision-making process (Black et al., 1999, p.240; Hutchings and Raine, 2006). However, the items and response formats generated was shaped, to an extent, according to the predispositions and biases of the decision-makers (Patton, 2002) – a potential limitation of the study. To offset this limitation, the items and response formats were reviewed in Study 3 by a larger group of experts in the fields of addiction and psychotherapy. Specifically, the inclusion of the Delphi exercise provided a complementary perspective, ensuring that the BATS was developed as comprehensively as possible (Morse, 2003).

#### **4.7 Conclusion**

An item pool for the BATS was generated using the measures identified from the literature as a basis. Generation of the items was primarily based on the results of a thematic analysis (Braun and Clarke, 2006); items from the identified measures were grouped according to what aspect of therapeutic practice they targeted. Thirty-three themes were developed and grouped into five meta-themes<sup>65</sup>. The thematic structure was refined, in collaboration with the supervision team, to identify the themes most relevant to the BATS; the number of themes was reduced from 33 to 18. The remaining 18 themes were expressed as individual items. Most of these items were adapted from the extracted items on the identified measures; eight were newly constructed. The generated items reflected the key features of therapies widely used in the treatment of alcohol and drug use. At this stage, consideration was also given to scoring; two response formats were developed. The formats reflected existing methods of evaluating treatment adherence and therapist competence. The next chapter describes how experts in the fields of addiction and psychotherapy reached a consensus on the content of the BATS (i.e., the items and response format).

---

<sup>65</sup> The five meta-themes were: i) session management, ii) medication and case management, iii) interventions to increase awareness, iv) interventions to change behaviour, and v) core skills.

## **Chapter 5**

### **Agreeing the content**

#### **5.1 Introduction**

The previous chapter described the second of four studies undertaken to develop the BATS; Study 2 generated an item pool and two response formats for potential inclusion in the BATS using the identified measures as a basis. This chapter presents the next of the four studies. This third study generated a consensus among selected experts on the content of the BATS (i.e., the items and response format). This chapter describes how a consensus was reached using Delphi methodology. The findings from the Delphi were used to develop the BATS. The findings and the newly developed measure are presented in this chapter.

#### **5.2 Method**

A Delphi approach was used to reach a consensus from selected experts on the content of the BATS. The starting point of this study was the item pool and response formats generated in Study 2. Data were collected using a series of questionnaires, or rounds, in which participant responses from one round were used to inform the next round. The iterative process combined experts' knowledge and opinions to develop a group consensus of opinion (McKenna, 1994). The approach is based on the adage "*two heads are better than one*" (Dalkey, 1969, p.411); the assumption that group responses are more reliable and valid than individual opinions (Hasson et al., 2000). The approach is advantageous in situations where there is a lack of evidence on the topic; in this case, the contributory factors of therapeutic change. An expert group provided a more updated exchange of information than the researcher and supervision team (GL, BB, and GT) in deciding on the content of the BATS.

##### **5.2.1 Participant selection**

Purposeful sampling was used to recruit participants. Purposeful sampling is a technique for selecting "*information-rich cases*" within the pragmatic constraints of time and resources (Patton, 2002, p.230; Suri, 2011). Information-rich cases are individuals, or

experts, who are especially knowledgeable about the topic under investigation (Palinkas et al., 2015). Potential participants were selected for their expertise in the areas of addiction and psychotherapy. Expertise was not limited to addiction; individuals working in other clinical areas may have knowledge of psychological therapies widely used for alcohol and drug use problems.

Expertise was defined using all three criteria: i) more than ten years of experience working clinically with clients and/or conducting research in the areas of addiction and/or psychotherapy, ii) a professional or an academic qualification in a relevant discipline, such as clinical psychology, and iii) a track record of peer-reviewed publications. Experience of process rating was considered desirable, although not essential. The inclusion of experts without process rating experience enabled different perspectives to be considered, encouraging a wider range of suggestions for improving the content of the BATS (Murphy et al., 1998). Selecting clinicians and academics aimed to improve the real-world applicability of the BATS (Hsu and Sandford, 2007), and enhance credibility with the target users<sup>66</sup> (Powell, 2003).

Potential participants were identified by searching the literature for key authors, and through discussion with the supervision team; the researcher and supervision team recommended potential participants based on their knowledge of the topic. Using two recruitment strategies reduced bias, as not all participants were selected on the basis of their acquaintance with the researcher and supervision team (Murphy et al., 1998). Fourteen experts were identified from the literature, and 10 were identified by the researcher and supervision team. Inevitably there was some overlap; five of the experts identified from the literature were also personal recommendations. In total, a list of 19 experts was compiled from which to approach potential participants.

The list of 19 experts was large enough to allow for dropouts and non-responders, while ensuring the participants had expertise in the areas of addiction and psychotherapy (Donohoe and Needham, 2008). A sample size of 10-15 participants was preferred (Delbecq et al., 1975); the reliability of group judgements declines rapidly with sample sizes of six or less, while above about 12, "*improvements in reliability is subject to diminishing returns*" (Murphy et al., 1998, p.37).

---

<sup>66</sup> Target users of the BATS are therapists, trainers, and practice supervisors.

It should be noted that snowball sampling was initially used as a third recruitment strategy. In round one, participants were invited to nominate other experts who might like to take part in the study. Twelve individuals were nominated. The nominees were not contacted because: i) the response rate was in line with the preferred sample size, ii) the nominees did not add additional diversity to the group; for example, they aligned to a similar therapeutic approach as the participants, iii) up-to-date contact details could not be obtained for two individuals, and iv) a third nominee had already been invited to take part in the study.

## **5.2.2 Measures**

Four questionnaires were developed for the study. The questionnaires were piloted with the supervision team and changes to the wording and formatting made. The questionnaires were completed anonymously. Anonymity encouraged the critique of ideas “*unbiased by the identities and pressures of others*” (Hasson et al., 2000 p.1012). An overview of the four questionnaires is provided below.

### **5.2.2.1 Demographic questionnaire**

The demographic questionnaire collected data on participants: gender (male, female, or transgender), place of residence (Asia, Africa, Australia, Europe, North America, or South America), length of experience in addiction and psychotherapy (years of experience in each area), and role (academic, clinician, dual role<sup>67</sup>).

### **5.2.2.2 Round one questionnaire**

The first round questionnaire contained 18 items for potential inclusion in the BATS. Items were generated from existing fidelity measures relevant to the addiction field. Participants were asked to consider each item and rate the extent to which they agreed: i) the item was important and should be included in the BATS, and ii) the item was comprehensible. Ratings were made on 7-point Likert-type scales (**Figure 6**). The response format featured a 7-point rating scale to provide scope for nuance in opinion. Also, the number of response options could be reduced to five or three levels, avoiding end-aversion bias<sup>68</sup> (Streiner et al., 2015; Hasson and Arnetz, 2005). The number was not higher than seven (e.g., a 9-point scale), because the literature suggests that “*people are unable to discriminate much*

---

<sup>67</sup> Dual role includes both clinical and academic components.

<sup>68</sup> End-aversion bias is the tendency for respondents to avoid rating the extreme ends of a scale.

beyond seven levels” (Streiner et al., 2015, p.48). Participants were invited to use free text spaces to provide comments about each item including suggestions for rewording.

	(1) strongly disagree	(2) disagree	(3) disagree somewhat	(4) neither disagree or agree	(5) agree somewhat	(6) agree	(7) strongly agree
The item is important	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The item is comprehensible	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Figure 6: Screenshot from the first round questionnaire of the 7-point scales used to rate item importance and comprehensibility**

After rating the items, participants were invited to share: i) their experiences of supervising staff, including use of supervision protocols and/or therapist rating scales, and ii) any additional comments about the items, scale or study. Participants’ experiences of supervising staff provided a means of cross checking the suitability of participants for inclusion in the study. Participants were invited to nominate other experts who might like to take part in the study. Appendix D provides an example screenshot from the first round questionnaire.

### **5.2.2.3 Round two questionnaire**

In round two, participants were asked to indicate whether they had completed the first round questionnaire. This enabled the consistency of participation to be assessed across rounds. Participants were invited to consider 14 items, developed by the researcher and the supervision team based on information provided in the first round. The results of round one were not summarised and fed back to participants in round two. Omitting the summarised data minimised any pressure to conform with group ratings (Hsu and Sandford, 2007). Participants were asked to rate the extent to which they agreed: i) the item was important and should be included in the BATS, and ii) the item was comprehensible. Ratings were made on 7-point scales (**Figure 6**). Participants were able to comment on each item, including suggestions for rewording, using free text spaces.

At this stage, participants were asked to consider how items on the BATS should be scored. Two formats were proposed: extensiveness, and quality. Example items along with the proposed rating scales were presented (**Figure 7**). For each of the proposed formats, participants rated the extent to which they agreed the format was an appropriate rating scale for inclusion in the BATS. Participants’ ratings were made on 7-point scales. Free text

space enabled participants to: i) comment on the proposed formats, such as, the number of anchors, and the proposed labels, and ii) suggest alternative rating formats. Lastly, free text space was given for participants to share additional comments about the items, scale or the study.

**1. Extensiveness** ("To what extent...")

"To what extent did the therapist explore both the positive and negative aspects of the client's substance use?"

<b>0</b> NOT AT ALL	<b>1</b> A LITTLE	<b>2</b> SOMEWHAT	<b>3</b> CONSIDERABLY	<b>4</b> EXTENSIVELY
------------------------	----------------------	----------------------	--------------------------	-------------------------

**2. Quality** ("How well...")

"How well did the therapist explore both the positive and negative aspects of the client's substance use?"

<b>0</b> VERY POOR	<b>1</b> POOR	<b>2</b> GOOD ENOUGH	<b>3</b> WELL	<b>4</b> VERY WELL
-----------------------	------------------	-------------------------	------------------	-----------------------

**Figure 7: Screenshot from the second round questionnaire of the proposed response formats for the BATS**

#### **5.2.2.4 Round three questionnaire**

In round three, participants indicated which of the previous rounds they had completed, enabling consistency of participation to be assessed across rounds. Participants considered 12 items, developed by the researcher and supervision team based on feedback provided in the second round. The results of round two were summarised and fed back to participants in round three. For each item, participants were shown the percentages of individuals who agreed: i) the item was important, and ii) the item was comprehensible<sup>69</sup>. The median rating and interquartile range (IQR) were also given (**Figure 8**). The summarised data encouraged participants to consider their judgements in light of the group's opinion on each item.

---

<sup>69</sup> Percentage agreement calculated by dividing the number of participants in agreement (scores of 5 'somewhat agree', 6 'agree', and 7 'strongly agree') by the total number of participants.

Participants rated the extent to which they agreed the items were important and should be included in the scale. Ratings were made on 7-point scales<sup>70</sup>. The questionnaire was shortened to encourage a higher response rate, and to avoid repetition. Item comprehensibility was not rated; it was anticipated that information provided by previous iterations would be sufficient for amending item wording. Free text spaces were given for participants to provide any additional comments about the items, scale or the study.

### **Item 1: "To what extent did the therapist keep the session focused on the aims for that session?"**

---

#### **Round 2 summary of results:**

- Item 1 is **important**: 92% agreed (median: 6, IQR: 5,7).
- Item 1 is **comprehensible**: 92% agreed (median: 7, IQR: 6,7)

Ratings were made on a scale of 1-7, where 1 is "*strongly disagree*" and 7 is "*strongly agree*".

#### **Figure 8: Screenshot from the third round questionnaire of the summary of results from round two for item 1**

### **5.2.3 Procedure**

The Delphi survey was conducted in three rounds using Bristol Online Surveys (Jisc, 2018). It was anticipated that three rounds would be sufficient, as participants were not asked to generate the items. Additional rounds were not offered, reducing the potential for participant fatigue (Hasson et al., 2000; Powell, 2003). The first round took participants approximately 15-20 minutes to complete, the second and third rounds 10-15 minutes. Each round was structured as follows:

- Study information and consent.** Participants were provided with information about the study. Details on what the study would involve were specific to each round (Appendix D shows the information sheet used in round one). A separate consent page was included (Appendix D). Participants gave their informed consent to take part in each round; completion of an earlier round was not a prerequisite for participating in a later round. Participants consented by submitting the completed survey.

---

<sup>70</sup> Likert-type scales were similar to **Figure 6**, but only one row was included: 'the item is important'.



- ii) **Demographic questionnaire.** Because the questionnaires were anonymous, participants were asked to complete the demographic questionnaire in all three rounds.
  
- iii) **First, second, or third round questionnaire.** The questionnaires for rating the items were specific to each round. In round one, for example, participants were given the first round questionnaire.

In round one, potential participants ( $n=19$ ) were sent an email inviting them to take part in the study. The email explained the aims of the study and what participation would involve; a link to the relevant questionnaires was included. Email reminders were sent approximately two weeks later. Because the responses were anonymous, the reminder emails thanked those who had taken part, and encouraged those who had not yet participated to complete the survey before the closing date. Subsequent rounds followed a similar process. Each round began shortly after data analysis of the previous round. Participants were given 4-5 weeks to complete each round; data collection and analysis ran from November 2016 to March 2017.

## **5.2.4 Data analysis**

### **5.2.4.1 Participant data**

Participant response rates were calculated for each round. Participants' demographic information was summarised. Qualitative data on participants' experiences of supervising staff from the first round questionnaire were also summarised. This qualitative data provided a means of cross-checking the suitability of participants for inclusion in the study.

### **5.2.4.2 Item data**

Statistical summaries of participants' ratings were produced for each item. The summaries were used to compare items across rounds. In rounds one and two, data were analysed on two dimensions: importance, and comprehensibility. In round three, analyses focused solely on item importance. The following descriptive statistics were computed:

- i) **Median, IQR, and range:** The median represented the level of group agreement; a higher median rating was indicative of a higher level of group agreement. The IQR reflected the degree of group consensus; a lower IQR

indicated a higher degree of convergence of group opinions. The range was computed to show the degree of dissent or divergence among participants (Powell, 2003).

- ii) **Percentage agreements:** Percentage of participants scoring between: i) 5 'somewhat agree' and 7 'strongly agree', and ii) 6 'agree', and 7 'strongly agree'. The percentage scores enabled participants' ratings across rounds to be compared. An increase in percentage scores across rounds was indicative of a change in participants' views towards consensus (Holey et al., 2007). Using two calculations provided a better representation of how participants' judged the items.
  
- iii) **Number of comments.** The number of comments (both positive and negative) was totalled for each item. A decrease in comments across rounds supported a move towards group consensus (Holey et al., 2007).

Item-specific feedback in all three rounds was collated and reviewed. In rounds one and two, the feedback and statistical summaries were used by the researcher and supervision team to make a decision on whether to retain, revise, or remove items for use in subsequent rounds. To aid the decision process, frequency distributions of participants' scores were produced for each item. The graphs showed visually what participants thought of items in terms of importance and comprehensibility. The graphs were important for interpreting the results; for example, median ratings can be misleading when interpreting bimodal data (Nestor and Schutt, 2015). The scores were grouped: 1-3 disagreement, 4 neither disagreement or agreement, and 5-7 agreement. In round three, the frequency distributions focused solely on comprehensibility.

#### **5.2.4.3 Item scoring data**

The second round questionnaire asked participants to consider how items on the BATS should be scored. Data on two possible response formats were collected: extensiveness and quality. Data were analysed using descriptive statistics, similar to those outlined previously:

- i) **Median, IQR, and the range.**

- ii) **Percentage agreement:** percentage of participants scoring between: i) 5 'somewhat agree' and 7 'strongly agree', and ii) 6 'agree', and 7 'strongly agree'.

Participants were able to provide qualitative feedback on the proposed formats. The feedback and descriptive statistics were used to decide which of the two scoring methods to include in the BATS. One round was considered sufficient for obtaining the information needed on item scoring; the main aim of the study was to agree on the items for inclusion in the BATS.

#### **5.2.4.4 Consensus**

Delphi studies generate a consensus of opinion among experts using a series of questionnaires or rounds (Hasson et al., 2000). In the final round, convergence of opinions is usually shown, with dispersion of participants' views lessening over successive rounds (Powell, 2003; Delbecq et al., 1975). There are no universally agreed criteria for defining consensus (Holey et al., 2007). The criteria depends on the number of participants, the research aims, and the resources available (Hasson et al., 2000). Studies have adopted different criteria for defining consensus; for example, achieving a certain level of central tendency, or the proportion of scores falling within a particular range (Diamond et al., 2014).

Data collection and analysis was an iterative process. It was, therefore, difficult to predict how participants would view the items. This uncertainty was a consideration in establishing the definition of consensus. Consensus was defined *a priori* using two criteria: a median rating between 5 and 7<sup>71</sup>, and an IQR of 2 or less. A narrower IQR was not used, an IQR of 1 or less is considered appropriate for 4- or 5-point Likert-type scales (von der Gracht, 2012). For consensus to be achieved, participants' scores needed to: i) meet the central tendency and dispersion criteria, and ii) be consistent with the item-specific feedback. Item-specific feedback referred to the qualitative comments provided by participants about each item. Participants' scores needed to reflect their comments to achieve consensus (e.g., high scores and positive feedback). In cases where there was a discrepancy between participants' scores and the qualitative feedback (e.g., high scores and negative feedback), weight was given to participants' feedback. Participants'

---

<sup>71</sup> On a 7-point Likert scale: 5 = Agree somewhat, and 7 = Strongly agree.

comments provided context for the judgements made.

A specific threshold was not specified for the range, percentage agreement scores (scoring between 5 and 7), and number of qualitative comments. These statistics were designed to complement the measures of central tendency and dispersion. For example, the range enabled differences in opinion among participants to be explored.

#### **5.2.4.5 General feedback**

Participants were invited to share additional comments about the items, scale, and study in all three rounds. In rounds one and two, the feedback was used to inform subsequent rounds; for example, missing items considered relevant for inclusion. In round three, the feedback helped to inform the first version of the BATS, for example, the scale structure.

### **5.3 Results**

#### **5.3.1 Participants**

Participation across rounds is summarised in **Table 16**. Nineteen experts were invited to take part in round one. One person declined by email, stating they did not have sufficient knowledge of the topic. Eighteen experts were approached for rounds two and three. The numbers of participants who took part in each round were: 12 (63%) in round one, 12 (67%) in round two, and 10 (56%) in round three. These numbers were in line with the preferred sample size of 10-15 individuals (Delbecq et al., 1975; Murphy et al., 1998).

**Table 16: Participation across rounds**

Participants	Round One	Round Two	Round Three
Number approached	19	18	18
Number of respondents (%)	12 (63%)	12 (67%)	10 (56%)
Completed one round	3	1	0
Completed two rounds	1	3	2
Completed three rounds	8	8	8

Participation across rounds was assessed by asking participants to indicate which rounds they had completed previously. There was a discrepancy between the number of

participants who completed round two and the number of participants who recalled completing the round. **Table 17** shows participants' self-reported participation across rounds. In round two, 12 participants reported completing round one. In round three, 10 participants recalled completing round two. Of these 10 participants, nine reported they had also taken part in round one. This gives a total of 13 individuals who recalled completing the second round, a discrepancy as only 12 participants took part.

**Table 17: Participants' self-reported participation across rounds**

Participants	Round One (n=12)	Round Two (n=12)	Round Three (n=10)
Completed one round	0	0	0
Completed two rounds	3	4*	1
Completed three rounds	9	9	9
<b>Total</b>	<b>12</b>	<b>13</b>	<b>10</b>

\*4 = 3 participants who completed round two and recalled completing round one, and 1 participant who completed round three and recalled completing round two (and not round one).

The participants' demographic information was, therefore, used to gauge individual participation across rounds (Appendix D). Participation appeared fairly consistent. Nine participants in round two had completed round one. Ten participants in round three had completed round two. Of these 10 participants, eight had also taken part in round one.

For each round, participants completed a demographic questionnaire. Data gathered are summarised in **Table 18**. In round one, the majority of participants were male. The opposite was true in rounds two and three, which had a higher proportion of females. Across all rounds, at least 70% of participants resided in Europe. There was representation from North America, reflecting a more international perspective. The participants were highly experienced. Participants had over 15 years of experience in the areas of addiction and/or psychotherapy. Most participants had a dual role, with both clinical and academic components. The diversity of participants' backgrounds assured a wide base of knowledge and expertise (Powell, 2003).

In the first round questionnaire, participants were asked to share their experiences of supervising staff, including use of supervision scales and/or therapist rating scales. Six participants summarised their experiences of:

- i) Supervision (n=3): “Extensive experience of supervision in various models of therapy. No experience of using rating scales for that purpose” (Participant 3r1),
- ii) Therapist rating scales (n=1): “Involvement in [a clinical trial]” (Participant 7r1).
- iii) Supervision and therapist rating scales (n=2): “I have extensive experience in both areas” (Participant 1r1).

**Table 18: Participant characteristics across all three rounds**

Participant Characteristics			Round One (n=12)	Round Two (n=12)	Round Three (n=10)
Gender	Female	N (%)	4 (33%)	7 (58%)	7 (70%)
	Male	N (%)	8 (67%)	5 (42%)	3 (30%)
Residence	Europe	N (%)	9 (75%)	9 (75%)	7 (70%)
	North America	N (%)	3 (25%)	3 (25%)	3 (30%)
Years of experience	Addiction	Mean year (SD), min-max	32.5 (3.6) 27-38	22.8 (11.8) 0-42	26.20 (11.3) 4-40
	Psychotherapy	Mean year (SD), min-max	27.8 (10.4), 10-40	25.3 (11.7), 0-44	26.2 (14.7), 0-50
Role	Academic	N (%)	2 (17%)	1 (8%)	1 (10%)
	Clinician	N (%)	2 (17%)	4 (33%)	4 (40%)
	Dual role	N (%)	8 (67%)	7 (58%)	5 (50%)

% = Percentage; Dual role = Respondents whose role had academic and clinical components; Min-max = Minimum and maximum values; N = Number of participants; SD = Standard deviation; Years of experience = Number of years spent working in and/or researching in the areas of addiction and psychotherapy.

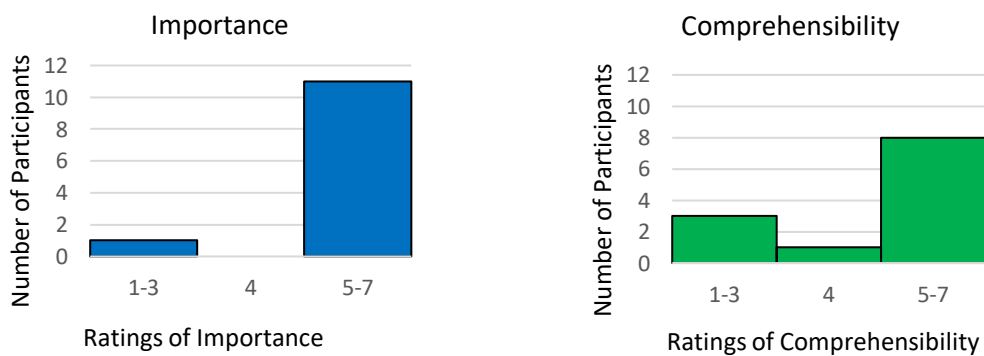
### 5.3.2 Round one

#### 5.3.2.1 Item data

Participants rated 18 items for potential inclusion in the BATS. Items were rated on two dimensions: importance and comprehensibility. The aim of the first round was to reduce the number of items, and to improve the comprehensibility of items for inclusion in the second round. Statistical summaries of participants’ ratings were produced for each item. The items are presented in **Table 19**. The table includes the median (representing the level of group agreement) and IQR (reflecting the degree of group consensus) for participants’ ratings of importance and comprehensibility across items.

In general, participants agreed that the items were important for inclusion in the BATS, but considered the wording unclear. Median ratings across items ranged from 5 to 7 for importance, and 3.5 to 7 for comprehensibility. The IQR showed minimal convergence of participants' opinions, particularly for item comprehensibility. The IQR across items was 1 to 3 for importance, and 1 to 5 for comprehensibility. Percentage agreement scores (participants scoring between 5 and 7) reflected the median ratings; higher percentages were indicative of higher median scores. Across items, the average range was 3.4 for importance, and 4.5 for comprehensibility, highlighting the differences of opinion. Appendix D summarises the statistical analyses for participants' ratings in round one.

The decision on whether to retain, revise, or remove items was made on a case-by-case basis. To aid the decision process, frequency distributions of participants' scores were produced for each item. The graphs showed visually what participants thought to items in terms of importance and comprehensibility. **Figure 9** shows the distributions of scores for item 1 'problem focused'. Item specific feedback was collated and reviewed. Participants' comments highlighted problems about the items and included suggestions for rewording. The statistical summaries and item-specific feedback were used to decide whether to retain, revise, or remove items. Proposed changes to the item pool were brought to the supervision team for discussion.



**Figure 9: Frequency distributions of participants' importance and comprehensibility ratings for item 1 'problem focused' in round one**

The researcher and supervision team agreed on what items to include for the second round. Item revisions were guided by the item-specific feedback. For example, **Table 20** shows the comments provided by participants for item 1 'problem focused'. Based on this information, the item was revised to: i) focus on one component of behaviour, ii) enable

**Table 19: Median and interquartile range (IQR) for participants' ratings of item importance and comprehensibility in round one**

Item Reference	Item	Importance		Comprehensibility	
		Median	IQR	Median	IQR
1. Problem focused	To what extent did the therapist organise the session so that defined tasks were covered?	6	1	6	2
2. Exploring behaviours	To what extent did the therapist encourage the client to talk about their current behaviour or status quo?	5	3	4.5	3
3. Exploring substance use	To what extent did the therapist gather information on the client's past or recent use of alcohol and drugs (e.g. patterns of use, extent of urges/thought, extent of reduction in use, results of recent urine/breath tests)?	6	2	6	3
4. Exploring impact of substance use	To what extent did the therapist explore the positive and negative aspects of the client's substance use?	6	1	6	1
5. Exploring pros and cons of change	To what extent did the therapist cover both the pros and cons for change?	6	2	5	1
6. Developing discrepancy	To what extent did the therapist encourage the client to contrast their addictive behaviour with personal goals or values?	6	1	5.5	1
7. Talking about change	To what extent did the therapist encourage the client to talk about the reasons to change their substance use behaviour?	7	2	6	2
8. Treatment goals	To what extent did the therapist develop or review the client's goals for treatment?	6.5	1	7	2
9. Behaviour change planning	To what extent did the therapist develop or review a change plan (e.g. steps, possible obstacles, sources of support, and solutions)?	7	1	6	2



Item Reference	Item	Importance		Comprehensibility	
		Median	IQR	Median	IQR
10. Identifying sources of support	To what extent did the therapist discuss the client's social network, exploring the nature of the client's relationships with network members and how those relationships might be used to support recovery?	6	2	5.5	2
11. Involving others	To what extent did the therapist discuss the availability of specific individuals who will be or are sources of support for the client's involvement in treatment or efforts to change their substance use behaviour?	6	2	6	2
12. Homework assigned	To what extent did the therapist with the client to plan specific tasks for the client to engage in between sessions?	6	2	3.5	4
13. Homework reviewed	To what extent did the therapist review previously assigned homework with the client?	7	1	7	1
14. Strengths and affirmations	To what extent did the therapist focus on the client's strengths, abilities, or efforts to change?	7	1	6	2
15. Collaboration	To what extent did the therapist actively attempt to engage the client in working together to explore therapeutic issues?	6	2	3.5	5
16. Empathy	To what extent did the therapist communicate understanding of and sensitivity to the client's comments and concerns?	7	1	6.5	1
17. Simple reflections	To what extent did the therapist use simple reflections by repeating (exact words or with slight rewording) what the client had said?	5	3	6	3
18. Complex reflections	To what extent did the therapist use complex reflections by paraphrasing what the client had said (i.e. adding new meaning and enabling the client to make connections)?	6	2	5	2

IQR = Interquartile range.

intra-session flexibility, i.e., reference to explicit agenda setting was omitted, and iii) reduce ambiguity, 'defined tasks' was replaced by 'the aims'. The changes to item 'problem focused' can be seen below:

- i) Original item included in round one: *"To what extent did the therapist organise the session so that defined tasks were covered?"*
- ii) Revised item included in round two: *"To what extent did the therapist keep the session focused on the aims for that session?"*

**Table 20: Participants' comments, and ratings of importance and comprehensibility for item 1 'problem focused' in round one**

Participant	Ratings		Comments
	Imp	Comp	
P4r1	4	5	<i>"I think I am old school and want to retain some intra session flexibility to deal with the unexpected."</i>
P5r1	6	4	<i>"Could fall out either on establishing an agenda or sticking to it. Try to make sure each question has only one component, otherwise gets too difficult."</i>
P7r1	6	5	<i>"A bit complicated? Does it mean tasks defined in a protocol or defined at the beginning of the session? Defined by whom?"</i>
P9r1	7	7	<i>"I initially read this item as having two parts. There is the organising the session which is important and I would take to mean covering such things as recap, goal setting, and some planning. The defined tasks suggest something more, such as practising a particular skill."</i>
P11r1	6	6	<i>"Item does not reflect importance of collaborative nature of agreement between patient and therapist."</i>

Imp = Importance; Comp = Comprehensibility

In some cases, there was a discrepancy between the statistical summaries and item-specific feedback. For example, three participants' highlighted the potential redundancy of item 4 'exploring impact of substance use', but the statistical summary supported the decision to retain the item. In such cases, items were revised to improve comprehensibility for the second round. In round one, no items were retained unchanged, 14 were revised, and four were removed.

**Table 21** summarises the items removed and the reasons for the decision. For example, item 2 'exploring behaviours' provided limited evidence for convergence of group opinions; the IQR was 3 for both importance and comprehensibility. The range was

similarly wide: 3-7 for importance, and 2-7 for comprehensibility. Participants' comments supported the decision to remove the item:

*"Ungrammatical again! And what does it mean? Does the 'or' signify alternatives? What type of behaviour is meant: substance use or something else? What on earth does status quo mean?" (Participant 7r2).*

**Table 21: Items removed in round one**

Item reference	Reasons for removal
2. Exploring behaviours	Limited evidence of group agreement on item importance. Ratings for item comprehensibility were relatively poor. Overlap with items 4 'exploring impact of substance use', and 5 'exploring pros and cons of change'.
7. Talking about change	Item similar to item 5 'exploring pros and cons of change'. Participants highlighted the redundancy of item 7, supporting the inclusion of item 5.
11. Involving others	Item not sufficiently distinct from item 10 'sources of support'. Statistical summaries supported item 10, rather than item 11.
17. Simple reflections	No evidence of group consensus for item importance or comprehensibility. Item considered therapy specific, not transtheoretical. Captured by item 16 'empathy'.

### **5.3.2.2 General feedback**

Eight participants provided additional comments about the items and the scale. The comments were generally positive and provided direction on item wording. Three participants emphasised the need for items to be clear and non-redundant: *"Be careful of complex/multifaceted items – they need to be very clear to aid reliability"* (Participant 11r1). One participant highlighted the importance of avoiding therapy specific jargon, particularly motivational interviewing. Three participants made suggestions for including items that focused on: i) *"being positive and hopeful about the future"* (Participant 8r1), ii) *"promotion of 12-step or other recovery-orientated self-help program participation"* (Participant 1r1), iii) *"other issues including serious life events or abnormal mental states"* (Participant 2r1). One participant advised giving more thought to the selection of items:

*"Involvement of the social network would be crucial and universal and seems under-represented. There needs to be some more thought on what is universal - I think support systems, having clear goals and getting people to do homework tasks stand out as the key elements of any intervention. All interventions should be delivered in a motivational style." (Participant 9r1)*

The feedback was used to inform round two, particularly regarding item relevancy and comprehensibility.

### **5.3.3 Round two**

#### **5.3.3.1 Item data**

Participants rated 14 items, developed by the supervision team based on information provided in the first round. Items were rated on two dimensions: importance and comprehensibility. The aims of round two were to build a consensus on what items to include in BATS, and to improve the comprehensibility of items for the third round. Statistical summaries of participants' ratings were produced for each item. The items are presented in **Table 22**. The table includes the median (representing the level of group agreement) and IQR (reflecting the degree of group consensus) for participants' ratings of importance and comprehensibility across items.

Participants were in agreement that the items were important and should be included in BATS. Median importance ratings ranged from 5.5 to 7, similar to the first round. Participants agreed the items were comprehensible. Median comprehensibility ratings ranged from 6 to 7, an improvement on the first round. Item 9 'homework assigned' best illustrates this improvement: the median comprehensibility rating was 3.5 in round one, and 7 in round two.

The IQR across items provided evidence of group consensus for importance and comprehensibility. The IQR for item importance ranged from 0 to 2. The IQR was lower when compared to the first round: the IQR ranged from 1 to 3 in round one. The IQR for item comprehensibility ranged from 1 to 2 with the exception of item 14 'complex reflections' (IQR: 4). In some cases, the IQR for item comprehensibility was wider than in the first round. For example, the comprehensibility IQR for item 3 'exploring impact of substance' was 1 in round one, and 2 in round two. New participants (those who completed round two but did not respond to the first round of questionnaires) may explain the wider spread of opinions.

Percentage agreement generally reflected the median scores across items. A higher percentage of participants who scored between 6 and 7 was indicative of a lower IQR. The average range across items was 2.6 for importance, and 3.1 for comprehensibility. The latter can be explained, in part, by the degree of dissent shown for item 14 'complex

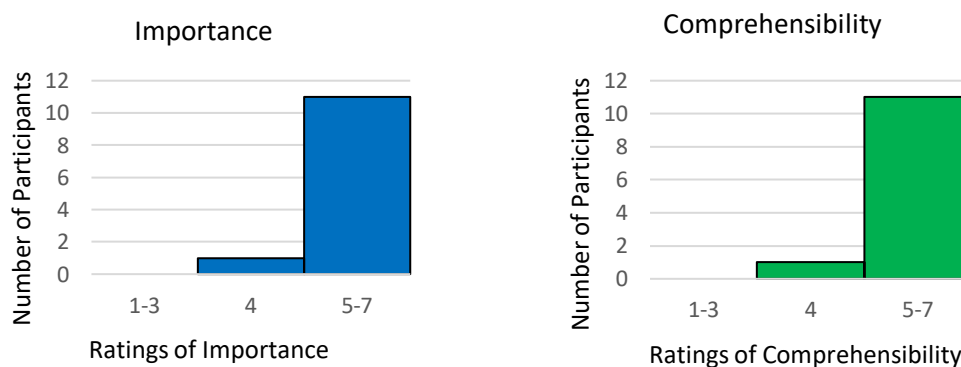
**Table 22: Median and interquartile range (IQR) for participants' ratings of item importance and comprehensibility in round two**

Item Reference	Item	Importance		Comprehensibility	
		Median	IQR	Median	IQR
1. Problem focused	To what extent did the therapist keep the session focused on the aims for that session?	6	2	7	1
2. Exploring substance use	To what extent did the therapist gather information on the client's history of substance use, or use since the last session?	5.5	2	7	1
3. Exploring impact of substance use	To what extent did the therapist explore both the positive and negative aspects of the client's substance use?	6	2	7	2
4. Exploring pros and cons of change	To what extent did the therapist encourage the client to talk about both the positive and negative aspects of changing substance use?	7	2	7	2
5. Developing discrepancy	To what extent did the therapist encourage the client to consider inconsistencies between their substance use, and personal goals or values?	6	1	6.5	1
6. Treatment goals	To what extent did the therapist develop and/or review the client's goals for treatment?	7	2	7	1
7. Behaviour change planning	To what extent did the therapist develop and/or review plan for changing the client's substance use?	6	2	7	1
8. Sources of support	To what extent did the therapist discuss how the client's relationships might be used to support changing substance use?	6	2	6	2
9. Homework assigned	To what extent did the therapist work with the client to plan tasks for the client to do between sessions?	7	1	7	1
10. Homework reviewed	To what extent did the therapist and the client review tasks planned in the previous session?	6	1	7	1
11. Strengths and affirmations	To what extent did the therapist focus on the client's strengths?	7	1	6.5	2
12. Collaboration	To what extent did the therapist attempt to work together with the client?	7	0	6.5	2
13. Empathy	To what extent did the therapist convey empathy?	7	0	7	1
14. Complex reflections	To what extent did the therapist use "complex reflections" – offering an interpretation, which adds new meaning and enables the client to make connections?	6	2	6	4

IQR = Interquartile range.

reflections’. The average range for both importance and comprehensibility was narrower than in the first round. Appendix D summarises the statistical analyses of participants’ ratings across items in round two.

As in round one, decisions on whether to retain, revise, or remove items were made on a case-by-case basis. Frequency distributions of participants’ scores showed visually what participants thought to items in terms of importance and comprehensibility. **Figure 10** shows the distributions of scores for item 1 ‘problem focused’; a shift towards agreement can be seen for both importance and comprehensibility when compared to the first round. The statistical summaries and item-specific feedback were used to inform the decision making process. Proposed changes to the items were brought to the supervision team for discussion. The supervision team decided on items for inclusion in the third round. In round two, five items were retained unchanged, seven were revised, and two were removed.



**Figure 10: Frequency distributions of participants’ importance and comprehensibility ratings for item 1 ‘problem focused’ in round two**

A consensus was reached for item comprehensibility with the exception of item 14 ‘complex reflections’. Feedback on how to improve the comprehensibility of item 14 was given. Participants highlighted two issues: First, the item would be difficult to rate, particularly for therapists unfamiliar with motivational interviewing. Item guidance was needed to add clarity ( $n=4$ ): “*You might need to offer an extended definition of complex reflections, e.g., via a link or footnote*” (Participant 11r2). Second, the inclusion of the word ‘interpretation’ was problematic ( $n=1$ ): “*Interpretations imply a theoretical framework (psychodynamic) which is not necessarily true for offering a complex reflection*” (Participant 1r2).

**Table 23** summarises the items removed in round two and the reasons for the decision. For example, the decision to remove item 2 ‘exploring substance use’ was supported by the participants’ comments. The behaviour was not considered relevant for every session, as a substance use history is mainly sought in the first session. In some cases, the item was considered counterproductive:

*“The therapist does not necessarily need to ask direct questions about this – a therapist behavior which could be counter-productive - but rather elicit the information if needed by asking other types of open questions that would naturally lead to such spontaneous disclosures on the client's part. At the same time, it could also be appropriate to make direct queries, depending on the context of the treatment and any existing agreements between client and therapist regarding monitoring of substance use.” (Participant 10r2)*

**Table 23: Items removed in round two**

Item reference	Reasons for removal
2. Exploring substance use	Participants’ comments did not support inclusion of the item, e.g. not applicable to all sessions, mainly only the first. Strong evidence of group agreement on item importance not shown. Overlap with item 4 ‘exploring pros and cons of change’, and item 14 ‘complex reflections’.
3. Exploring impact of substance use	Participants’ comments highlighted that too much dialogue on the use of substances can be counterproductive, particularly when exploring the positive aspects. Captured by item 4 ‘exploring pros and cons of change’, this had a higher median importance score.

Item 2 did not provide strong evidence of group agreement; the median importance score was 5.5. The item overlapped with item 4 ‘exploring pros and cons of change’, and item 14 ‘complex reflections’.

### **5.3.3.2 Item scoring data**

In round two, participants were asked to consider how items included in the BATS should be rated. Two response formats were proposed: extensiveness, and quality. Participants rated the extent to which they agreed these formats were appropriate for use in the BATS. Descriptive statistics summarised participants’ responses. Ratings were similar for both formats. For example, the median and IQR were 6 and 1 for extensiveness, and 6.5 and 1 for quality. The extensiveness scale was chosen based on participants’ comments ( $n=6$ ):

*“Extensiveness will be far easier to rate and more reliable. Quality requires judgements of competence and this is way more difficult and I suspect you will get poorer reliability.” (Participant 11r2).*

Five participants gave suggestions on how to improve the response format; for example, simplifying extensiveness to ‘how much’, and including a ‘not relevant/appropriate for this session’ response option.

### **5.3.3.3 General feedback**

One participant provided additional feedback: *“I have made comments along the way. Hope they are helpful”* (Participant 12r2).

## **5.3.4 Round three**

### **5.4.3.1 Item data**

Participants rated 12 items, developed by the supervision team based on information provided in the second round. Unlike previous rounds, participants rated item importance only. Item comprehensibility was not rated, information provided by previous rounds was considered sufficient for amending item wording. The aim of round three was to reach a consensus on what items to include in the BATS. Statistical summaries of participants’ ratings were produced for each item. The items are presented in **Table 24**. The table includes the percentage agreements, median, IQR, and range for participants’ ratings of importance across items. **Table 25** summarises participants’ ratings of the items over successive rounds.

Median ratings show a high level of group agreement on the importance of the items for inclusion in the BATS. Median scores ranged from 6.5 to 7, and were generally higher than in previous rounds. The IQR provided evidence of group consensus. The IQR for most items was 0 to 1; three items had an IQR of 2. The IQR was generally narrower across items than in previous rounds. Percentage agreements (scoring between 5 and 7) were all above 80%, with seven items attaining 100% agreement. Compared to the second round, an increase in percentage agreements was shown for most items. An increase in percentage scores across rounds was less apparent when using the stricter definition of agreement (scoring between 6 and 7). The range across items varied between 0 and 4. The average range decreased over successive rounds: 3.4 in round one, 2.6 in round two, and 2.1 in round three.



**Table 24: Percentage agreements, median, interquartile range (IQR), and range for participants' ratings of item importance in round three**

Item Reference	Item	Importance Ratings		
		% Agreement (% Scoring 6-7)	Median (IQR)	Range
1. Problem focused	To what extent did the therapist keep the session focused on the aims for that session?	100% (100%)	6.5 (1)	6-7
2. Exploring pros and cons of change	To what extent did the therapist encourage the client to talk about both the positive and negative aspects of changing substance use?	100% (70%)	6 (1)	5-7
3. Developing discrepancy	To what extent did the therapist encourage the client to consider inconsistencies between their substance use, and personal goals or values?	100% (90%)	7 (1)	5-7
4. Treatment goals	To what extent did the therapist discuss the client's goals for treatment?	100% (100%)	7 (0)	6-7
5. Behaviour change planning	To what extent did the therapist discuss a plan for changing the client's substance use?	90% (80%)	6 (1)	5-7
6. Sources of support	To what extent did the therapist discuss how the client's social network might support changing substance use?	80% (60%)	6.5 (2)	4-7
7. Homework assigned	To what extent did the therapist and the client plan tasks for the client to do between sessions?	80% (60%)	6 (2)	3-7
8. Homework reviewed	To what extent did the therapist and the client review tasks planned in the previous session?	80% (70%)	6.5 (2)	3-7
9. Strengths and affirmation	To what extent did the therapist focus on the client's strengths?	100% (100%)	7 (0)	6-7
10. Collaboration	To what extent did the therapist attempt to work together with the client?	100% (100%)	7 (0)	6-7
11. Empathy	To what extent did the therapist convey empathy?	100% (100%)	7 (0)	7-7
12. Complex reflections	To what extent did the therapist use "complex reflections" – offering a perspective which adds new meaning and enables the client to make connections?	90% (90%)	6.5 (1)	4-7

% Agreement = Percentage of participants scoring between 5 and 7; % Scoring 6-7 = Percentage of participants scoring between 6 and 7; IQR = Interquartile range.

**Table 25: Statistical summaries of participants' ratings for importance and comprehensibility across rounds**

Item Reference	Round	Importance Ratings			Comprehensibility Ratings			General
		% Agreement (% scoring 6-7)	Median (IQR)	Range	% Agreement (% scoring 6-7)	Median (IQR)	Range	Number of comments
1. Problem focused	1	92% (83%)	6 (6,7)	3-7	92% (67%)	6 (5,6.75)	2-7	5
	2	92% (67%)	6 (5,7)	4-7	92% (92%)	7 (6,7)	4-7	0
	3	100% (100%)	6.5 (6,7)	6-7	-	-	-	5
2. Exploring pros and cons of change	1	92% (75%)	6 (5.25,7)	4-7	83% (33%)	5 (5,6)	2-7	8
	2	92% (75%)	7 (5.25,7)	3-7	83% (75%)	7 (5.25,7)	3-7	2
	3	100% (70%)	6 (5,6.25)	5-7	-	-	-	2
3. Developing discrepancy	1	100% (92%)	6 (6,7)	5-7	92% (50%)	5.5 (5,6)	3-7	5
	2	100% (83%)	6 (6,7)	5-7	100% (83%)	6.5 (6,7)	5-7	0
	3	100% (90%)	7 (6,7)	5-7	-	-	-	1
4. Treatment goals	1	100% (92%)	6.5 (6,7)	5-7	83% (75%)	7 (5.25, 7)	2-7	3
	2	92% (75%)	7 (5.25,7)	4-7	83% (83%)	7 (6,7)	3-7	4
	3	100% (100%)	7 (6.75,7)	6-7	-	-	-	1
5. Behaviour change planning	1	92% (83%)	7 (6,7)	3-7	92% (58%)	6 (5,7)	3-7	4
	2	83% (75%)	6 (5.25,7)	4-7	92% (92%)	7 (6,7)	4-7	3
	3	90% (80%)	6 (5, 6.25)	5-7	-	-	-	4
6. Sources of support	1	92% (75%)	6 (5.25,7)	2-7	83% (50%)	6 (5.25,7)	2-7	8
	2	83 (67%)	6 (5,7)	4-7	83% (67%)	6 (5,7)	4-7	2
	3	80% (60%)	6.5 (4.75,7)	4-7	-	-	-	2
7. Homework assigned	1	83% (67%)	6 (5,7)	2-7	33% (33%)	3.5 (2,6)	1-7	10
	2	83% (83%)	7 (6,7)	4-7	100% (92%)	7 (6.25,7)	5-7	2
	3	80% (60%)	6 (4.75,7)	3-7	-	-	-	2

Item Reference	Round	Importance Ratings			Comprehensibility Ratings			General
		% Agreement (% scoring 6-7)	Median (IQR)	Range	% Agreement (% scoring 6-7)	Median (IQR)	Range	Number of comments
8. Homework reviewed	1	100% (83%)	7 (6,7)	5-7	92% (92%)	7 (6,7)	4-7	4
	2	92% (83%)	6 (6,7)	4-7	100% (92%)	7 (6,7)	5-7	2
	3	80% (70%)	6.5 (4.75,7)	3-7	-	-	-	1
9. Strengths and affirmations	1	92% (92%)	7 (6,7)	4-7	83% (58%)	6 (5,7)	3-7	4
	2	100% (83%)	7 (6,7)	5-7	92% (75%)	6.5 (5.25,7)	4-7	3
	3	100% (100%)	7 (7,7)	6-7	-	-	-	2
10. Collaboration	1	92% (58%)	6 (5,7)	2-7	50% (33%)	3.5 (2.25,6.75)	2-7	8
	2	100% (100%)	7 (7,7)	6-7	83% (75%)	6.5 (5.25,7)	3-7	6
	3	100% (100%)	7 (6.75,7)	6-7	-	-	-	1
11. Empathy	1	100% (100%)	7 (6.25,7)	6-7	83% (83%)	6.5 (6,7)	2-7	4
	2	100% (100%)	7 (7,7)	6-7	100% (100%)	7 (6.25,7)	6-7	1
	3	100% (100%)	7 (7,7)	7-7	-	-	-	1
12. Complex reflections	1	92% (75%)	6 (5.25,7)	2-7	75% (33%)	5 (4.25,6)	2-6	7
	2	100% (75%)	6 (5.25,7)	5-7	67% (58%)	6 (3.25,7)	1-7	5
	3	90% (90%)	6.5 (6,7)	4-7	-	-	-	3

% Agreement = Percentage of participants scoring between 5 and 7; % Scoring 6-7 = Percentage of participants scoring between 6 and 7; IQR = Interquartile range.

Item-specific feedback was provided. Feedback was grouped into four categories (**Table 26**). First, participants noted that 5 of the items were not always relevant for every session: *“Very important if ‘Tasks’ were set, but clearly irrelevant if not”* (Participant 2r3 commenting on item 5). Concerns about item relevancy may explain the wider IQR for three of the items. Second, participants provided suggestions for rewording for 5 of the items: *“Use of word session twice in one sentence, somehow feels clumsy and distracting”* (Participant 9r3 commenting on item 1). Third, the importance of item definitions was highlighted: *“Good, as long as it is clear what strengths is about”* (Participant 7r3 commenting on item 9). Lastly, participants gave general positive feedback for three of the items, validating their inclusion in the BATS: *“Essential, even when challenging interactions take place”* (Participant 2r3 commenting on item 11). The number of comments varied across items, and did not seem to relate to the median importance scores. There was a general reduction in comments as rounds progressed.

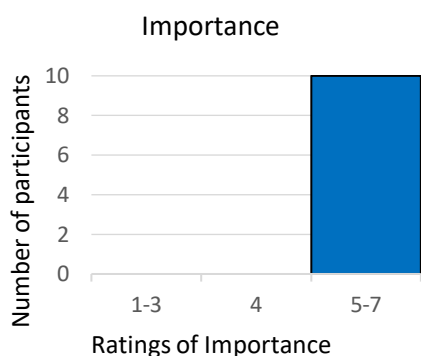
**Table 26: Item-specific feedback for items in round three**

Item Reference	Participants’ Comments				
	Item Relevancy	Item Wording	Item Guidance	General Feedback	Total
1. Problem focused	-	1	4	-	5
2. Exploring pros and cons of change	-	1	1	-	2
3. Developing discrepancy	1	-	-	-	1
4. Treatment goals	-	1	-	-	1
5. Behaviour change planning	3	-	1	-	4
6. Sources of support*	1	1	-	-	2
7. Homework assigned*	1	-	1	-	2
8. Homework reviewed*	1	-	-	-	1
9. Strengths and affirmation	-	-	1	1	2
10. Collaboration	-	-	-	1	1
11. Empathy	-	-	-	1	1
12. Complex reflections	-	3	-	-	3

\*Items had an Interquartile range (IQR) of 2 (all other items had an IQR of 1).

Decisions on what to do with the items were made on a case-by-case basis. Frequency distributions of participants’ scores showed visually what participants thought of the items. **Figure 11** shows the distributions of scores for item 1 ‘problem focused’; all participants agreed that item 1 was important and should be included in the BATS. The decision on

whether to retain, revise, or remove items was informed by the statistical summaries and item-specific feedback. Proposed changes to the item pool were brought to the supervision team for discussion. The supervision team decided on the items for inclusion in the BATS. Consideration was given to the order in which items were presented, the scale guidance, and the item definitions. In round three, nine items were retained unchanged, three were revised, and none were removed. These 12 items were included in the BATS.



**Figure 11: Frequency distributions of participants' importance ratings for item 1 'problem focused' in round three**

#### **5.4.3.2 General feedback**

Six participants provided additional feedback on the items, scale and the study.

Participants asked for clarity on the appropriateness of the BATS for targeting different stages of the therapy process: *"Main problem to solve is how to handle stage specific questions. The rater will not be aware of stage and maybe these need to be dropped or reworked"* (Participant 9r3). Clarification was asked for on: i) what the BATS would be rating, such as, video and audio recordings of treatment sessions, and ii) the aim of the BATS in relation to existing fidelity measures:

*"...Is the scale to be used based on video? Is it thought to be a complement to other types of in-session rating schemes like the MITI or CLAMI? Just some questions, hope they are helpful."* (Participant 5r3)<sup>72</sup>

The feedback was used to inform: i) item scoring, for example, scoring an item that was not appropriate or relevant to the session, and ii) instructions for use, including the purpose of

---

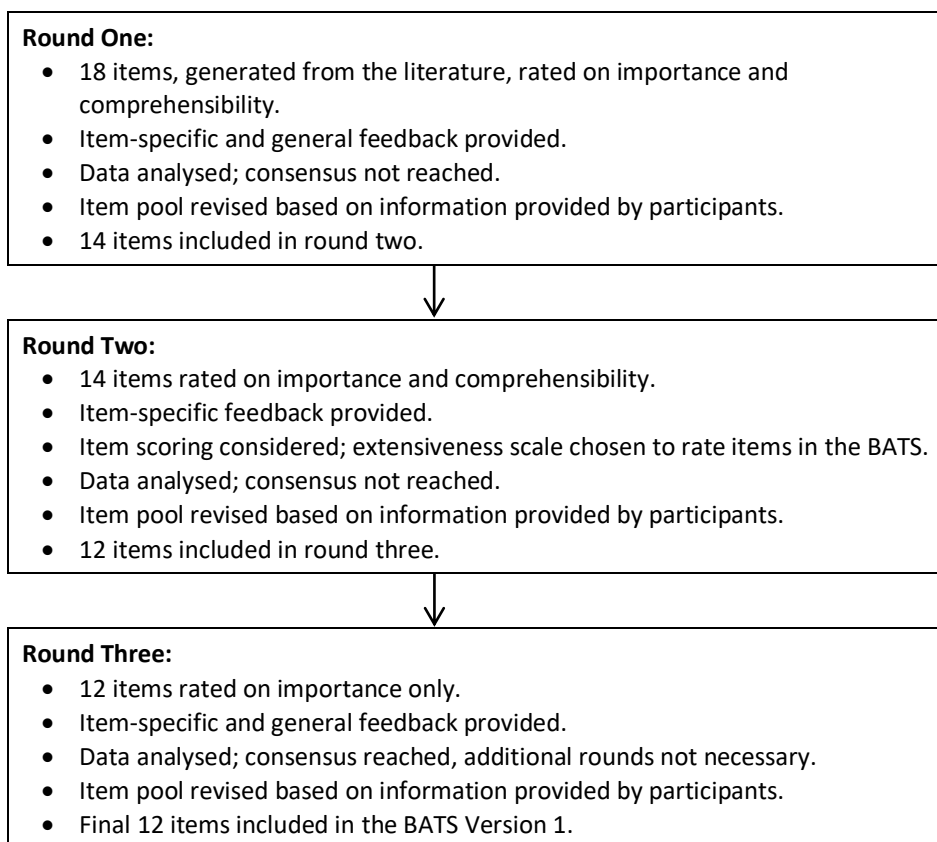
<sup>72</sup> MITI = Motivational Interviewing Treatment Integrity Code (Moyers et al., 2015); CLAMI = Client Language Assessment in Motivational Interviewing (Miller et al., 2008b).

the scale and how it should be used.

## 5.4 Discussion

### 5.4.1 Study overview

A Delphi study was used to obtain a consensus of opinion among selected experts on the content of the BATS. Participants were selected for their expertise in the areas of addiction and psychotherapy. The study involved three iterative rounds of data collection and analysis (**Figure 12**).



**Figure 12: Study 3 overview**

The 18-item pool generated in Study 2 was used as a starting point. Specifically, the first round aimed to reduce the number of items, and improve item comprehensibility. Participants rated the 18 items on two dimensions: importance and comprehensibility. Item-specific feedback was provided. The results were encouraging; 14 items were revised for inclusion in the second round questionnaire based on information provided by participants in round one. Round two aimed to reduce the level of dispersion among participants' views, and further improve item comprehensibility. Similar to round one,

items were rated on importance and comprehensibility. The item pool was reduced to 12 items based on information provided by participants; most items were revised to improve clarity. At this stage, consideration was given to how items included in the BATS should be scored. Two response formats were proposed: extensiveness, and quality. Participants rated the extent to which they agreed these methods of rating were appropriate for use in the BATS. Participants' ratings and written feedback were used to select one of the formats for inclusion in the BATS. Round three focused solely on item importance; data provided in round two was considered sufficient for amending item wording. The aim of round three was to reach a consensus on what items to include in the BATS. The results showed a group agreement on the items for inclusion in the BATS. The final list of 12 items and the chosen response format was compiled in the first version of the BATS (Appendix D).

## 5.4.2 Main findings

### 5.4.2.1 Item data

Consensus for each item was defined a priori using two criteria: a median rating between 5 and 7, and an IQR of 2 or less. For consensus to be achieved, participants' ratings had to be consistent with the item-specific feedback. **Table 27** summarises the number of items retained, revised, and removed across rounds.

**Table 27: Number of items retained, revised and removed across rounds**

	Items Retained	Items Revised	Items Removed	Total
Round One	0	14	4	18
Round Two	5	7	2	14
Round Three	9	3	0	12

In round two, participants rated 14 items on importance and comprehensibility. Thirteen items achieved consensus. Two items were removed because participants' ratings were not consistent with the item-specific feedback; the items were considered to be counterproductive in some instances. One item achieved consensus for importance only. Because participants judged the item important, it was not removed from the item pool but revised to improve clarity. In total, 12 items were included in the third round questionnaire: 5 items were retained (with no changes) and 7 were revised.

In the final round, participants rated 12 items on importance only. Item comprehensibility

was not rated, information provided in previous rounds was considered sufficient for amending item wording. All items achieved consensus. The median ratings showed a high level of group agreement, ranging from 6.5 to 7. The IQR provided evidence of convergence of participants' opinions. The IQR for most items was 0 to 1; three items had an IQR of 2. Item-specific feedback was generally positive. Feedback highlighted the importance of including item definitions and scoring instructions, particularly for items not considered relevant for every session. Suggestions for rewording items were given. In total, 9 items were retained unchanged, and 3 items were revised. These 12 items were included in the BATS.

#### **5.4.2.2 Item scoring data**

In round two, participants were asked to consider how items included in the BATS should be rated. Two scoring methods were proposed: extensiveness and quality. Participants' ratings were similar for both methods. The extensiveness scale was chosen, as participants' feedback suggested the extensiveness scale would be easier and more reliable to rate than the quality scale.

#### **5.4.2.3 General feedback**

Participants were able to leave general feedback in each round about the items, scale and the study. In round one, comments focused on item wording, for example, to amend items for clarity and to avoid therapy specific jargon. Suggestions for including additional items were also given. In round two, participants made no additional comments. In round three, the general feedback shifted from comments about item wording to queries on using the scale in practice. For example, clarification was needed on how therapy sessions would be rated, e.g. use of audio and video recordings, and how the BATS could evaluate therapy sessions targeting different stages of change.

### **5.4.3 Study strengths and limitations**

#### **5.4.3.1 Participants**

A key strength of this study is the participants. The participants were highly experienced, with over 15 years of experience in the fields of addiction and/or psychotherapy. Most participants had a dual role, both clinical and academic components. The diversity of participant backgrounds assured a wide base of knowledge and expertise (Powell, 2003). The selected experts also allowed for different perspectives to be considered (Murphy et



al., 1998), reducing the possibility of having overlooked some obvious facet in developing the BATS. Selecting clinicians and academics improves the real-world applicability of the BATS (Hsu and Sandford, 2007), and enhances credibility with the target users (Powell, 2003).

The numbers of participants who completed each round of the study were: 12 in round one, 12 in round two, and 10 in round three. The numbers were in line with the preferred sample size of 10-15 individuals, providing confidence in the reliability of the group judgements (Murphy et al., 1998). Participants were identified by searching the literature for key authors, and through discussion with the supervision team. Using two identification strategies reduced bias, not all participants were selected on the basis of their acquaintance with the supervision team (Murphy et al., 1998); a challenge given the specialised topic under investigation and the expertise of the supervision team. One potential weakness of this study is that snowball sampling, recommended in Delphi studies (Ludwig, 1997), was not used. This was not considered problematic, as participants who completed the first round questionnaire were considered sufficient in number to have provided "*a representative pool of judgements*" (Hasson et al., 2000, p.4).

Another potential weakness of this study is that the findings represent the views and opinions of a particular group of people at a particular point in time. This raises concerns over the reliability of the findings (Reid, 1993). Specifically, on whether the findings would be the same if a second, comparable group of experts were recruited. Replicability is difficult to show in Delphi studies, as participants are "*selected for a purpose, to apply their knowledge to a certain problem*" (Hasson et al., 2000, p.1010). There are examples in the Delphi literature of how replicability can be assessed. For example, Garnett et al. (2015) conducted a three-round Delphi study with seven experts to identify the aspects of treatment most likely to be effective for reducing alcohol consumption when delivered in a smartphone app. The experts generated a list of "*best bet*" intervention components and strategies, which were rated and ranked (Garnett et al., 2015, p.3). A second group of independent experts also ranked the identified components. Replicability was demonstrated by correlating the rankings made by the seven experts with those made by the independent expert group. In the current study, a second group of participants would have strengthened confidence in the findings. However, focus was given to recruiting a

larger pool of experts that were “*willing and able to make valid contribution*”, important for the study’s success in achieving group agreement on the content of the BATS (Powell, 2003, p.379).

#### **5.4.3.2 Questionnaires**

That the questionnaires were anonymous is considered a strength of the study. Participants were able to express their opinions without being influenced by each other’s responses (Murphy et al., 1998). The iterative process of data collection enabled participants to change their opinions without admitting they had done so to others in the group (Rowe and Wright, 1999). It also encouraged participants to make honest judgements based on their own experiential knowledge rather than adopting a more “*cautious institutional position*” (Gupta and Clarke, 1996, p.186).

That being said, there were three main limitations of using anonymous questionnaires. First, there was no accountability for the opinions expressed by participants, this may have encouraged participants to judge items without due consideration (Goodman, 1987). Second, anonymity may have lessened participants’ motivation to take part, their contribution was not known to others in the group (McKenna, 1994). Third, it may have increased participant fatigue. Because the responses were anonymous, participants were asked to complete the demographic questionnaire in all three rounds. Reminder emails were also sent to participants irrespective of whether they had already taken part. The first of these limitations was not considered problematic; the item-specific feedback provided by participants across rounds suggested that consideration was given to rating the items. However, it was difficult to gauge the impact of the second and third limitations. An alternative design would have been to use a ‘quasi-anonymous’ approach; participants are known to the researcher, in some cases to one another, but their judgements remain anonymous (McKenna, 1994). Knowing the identity of the participants would have enabled targeted follow-up with non-responders, and may have increased the response rates.

#### **5.4.3.3 Analyses**

There are “*no firm rules for establishing when consensus is reached*” (Powell, 2003, p.379). Similar to other Delphi studies, a consensus was reached in the final round, with the dispersion of participants’ views lessening over successive rounds. Consensus was reached when participants’ scores: i) met the central tendency and dispersion criteria, and ii) were

consistent with the item-specific feedback. It has been argued that inferential statistics are more appropriate and statistically sound than descriptive statistics, particularly for deciding when to terminate data collection (Kalaian and Kasim, 2012). Using inferential statistics may have increased rigour, as the results would have shown whether significant variation existed among the opinions of participants (Kalaian and Kasim, 2012). However, the lack of inferential statistics was not considered majorly problematic. Powell (2003) argues that Delphi studies should not be judged by the same validation criteria as “*hard science*” (p.380). Consensus development methods are not seeking to create new knowledge, they are processes for making the best use of available information (Murphy et al., 1998), in this case, the experiential knowledge of participants. As such, the complement of descriptive statistics together with the item-specific feedback was considered appropriate and sufficient for establishing consensus in this study.

## **5.5 Conclusion**

A consensus was generated among experts in the fields of addiction and psychotherapy on the content of the BATS (i.e., the items and response format). This study used a Delphi approach to obtain a consensus, and was based on the item pool and response formats generated in Study 2. Data were collected using three rounds of questionnaires, in which participant responses from one round were used to inform the next round. The iterative process combined experts’ knowledge and opinions to develop group agreement on the content of the BATS (McKenna, 1994). By the end of the third round, experts agreed that the revised 12 items were both comprehensible and important for inclusion in the BATS. Consideration was also given to item scoring. Two scoring methods were proposed: extensiveness and quality. The extensiveness scale was chosen for the BATS, as participants’ feedback suggested the extensiveness scale would be easier and more reliable to rate than the quality scale. Thus, this study led to the development of the first version of the BATS.

## **Chapter 6**

### **Testing reliability and validity**

#### **6.1 Introduction**

The previous chapter described the third of four studies undertaken to develop the BATS; a consensus was generated among selected experts in the fields of addiction and psychotherapy on the content of the BATS (i.e., the items and response format). This led to the development of the BATS version 1. This chapter details the fourth study, which aimed to test the reliability and validity of the newly developed measure. Specifically, the BATS was subjected to convergent validity, and inter-rater reliability analyses. A description of the data and analytic methods used in this study are presented. The findings are discussed in relation to the literature.

#### **6.2 Method**

The BATS was developed further by investigating its psychometric properties. Specifically, the BATS was subjected to convergent validity, and inter-rater reliability analyses. Convergent validity was examined in two ways. First, the relationships between the BATS and the Working Alliance Inventory (WAI) were explored. The Working alliance Inventory is a well-validated measure of therapeutic alliance. Therapeutic alliance refers to the collaborative relationship between the client and the therapist, and consists of three components: i) agreement over the goals of therapy, ii) agreement about the tasks of therapy – activities the client and therapist engage in as part of the therapeutic journey, and iii) the bond between the client and the therapist (Horvath and Greenberg, 1989). There is compelling evidence of the relationship between the alliance and the outcome of therapy (Wampold and Imel, 2015). The WAI and the BATS are both measures of treatment process. However, the BATS is a broader measure, designed to evaluate key features of therapies widely used in addiction, of which the alliance is one aspect. Therefore, it was predicted that there would be some overlap or association between the two measures.

Second, the BATS was compared with three existing fidelity measures: ADAPTA PRS (Tober and Crosby, 2014), AESOPS PRS (Tober and Crosby, 2011), and UKATT PRS (Middleton et

al., 2001)<sup>73</sup>. Inter-rater reliability analyses tested the consistency of measurement on the BATS between two different raters (HC & GT). The chosen analyses were informed by the methods used to validate the fidelity measures identified in Study 1. Data used in this investigation included: secondary analysis of data from three randomised controlled trials (RCTs), routine practice data, and process rating data derived from the BATS. A more detailed overview of the data and analyses are provided below.

### 6.2.1 Data sources

Four data sources were used in the development of the BATS (Table 28).

**Table 28: Data sources used in the development of the BATS**

Type	Source	Data
Secondary analysis of trial data	ADAPTA	Video recordings of 30-45 minute sessions of an alcohol focused (AF) and a healthy living (HL) intervention ( $n=50$ ). Process rating data and Working Alliance Inventory (WAI) data.
	AESOPS	Audio recordings of 5-minute sessions of brief advice (BA), and 20-minute sessions of behaviour change counselling (BCC) ( $n=160$ ). Process rating data.
	UKATT	Video recordings of 50-minute sessions of social behaviour and network therapy (SBNT) and motivational enhancement therapy (MET) ( $n=452$ ). Process rating data and WAI data.
Routine practice data	RP	Video recordings of therapy sessions delivered by therapists working at a NHS specialist addictions service (SAS) in the North of England ( $n=16$ ). Video recordings of therapy sessions delivered by therapists working at a non-NHS drug and alcohol service (DAS) in Wales ( $n=9$ ).

ADAPTA = Addressing Drinking Among Patients: comparing Two Approaches; DAS = Drug and alcohol service; AESOPS = Alcohol: Evaluating Stepped care in Older Populations Study; NHS = National Health Service; RP = Routine practice; SAS = Specialist addictions service; UKATT = United Kingdom Alcohol Treatment Trial.

### 6.2.2 Secondary analysis of trial data

Secondary analysis of data from three RCTs – ADAPTA (Watson et al., 2015), AESOPS (Watson et al., 2013a), and UKATT (UKATT Research Team, 2005) – comprised:

<sup>73</sup> ADAPTA PRS = ADAPTA (Addressing Drinking Among Patients: comparing Two Approaches); AESOPS PRS = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale; UKATT PRS = UKATT (United Kingdom Alcohol Treatment Trial) Process Rating Scale.

- i) Recordings of therapy sessions from the all three trials.
- ii) Process rating data from the trials.
- iii) Working Alliance Inventory (WAI) outcome data from ADAPTA and UKATT.

#### **6.2.2.1 ADAPTA**

ADAPTA was a pilot RCT, comparing an alcohol focused intervention (AF) and a healthy living intervention (HL) for adult problem drinkers identified in a general hospital setting (Watson et al., 2015). The AF intervention was a manual-based adaptation of social behaviour and network therapy (SBNT; Copello et al., 2002), which helped clients to build networks of people supportive of positive change in drinking (UKATT Research Team, 2001). In the HL intervention, participants could chose to change their behaviour in up to three health behaviour domains from a choice of seven: drinking, drug use, diet, smoking, exercise, personal care, and medication compliance (Watson et al., 2015). This intervention was based on the principles of behaviour change counselling (BCC). Both the AF and the HL interventions comprised four 30-45 minute sessions delivered by therapists working at a National Health Service (NHS) addictions service in the North of England.

In the trial, treatment sessions were recorded and rated to assess treatment fidelity. The sessions were recorded onto digital video disc (DVD) with visual recordings of therapists. Fifty sessions were rated as part of the trial, 42% of the total number of sessions ( $n=119$ ). Sessions were rated using the ADAPTA PRS<sup>74</sup> (Tober and Crosby, 2014). The 15-item fidelity measure covered the specific techniques described in the therapy manuals. Most items were scored on two 5-point scales, measuring the extent to which therapists carried out item behaviours (frequency), and how well therapists performed them (quality). Item scores ranged from: 0 (not at all/very poor) to 4 (extensively/very well). Data derived from the ADAPTA PRS were analysed to ensure treatment fidelity (Watson et al., 2015). The current study used this process rating data (derived from the ADAPTA PRS) to test the psychometric properties of the BATS.

In ADAPTA, a modified version of the Working Alliance Inventory (WAI; Horvath and Greenberg, 1989) was used to capture changes in therapeutic alliance (Watson et al., 2015). The 12-item short-form of the WAI (WAI-S; Tracey and Kokotovic, 1989) was

---

<sup>74</sup> ADAPTA PRS = ADAPTA (Addressing Drinking Among Patients: comparing Two Approaches) Process Rating Scale.

completed by both client and therapist at the end of the second and third sessions (Watson et al., 2013b). Items on the WAI-S were scored on 7-point scales, measuring the extent to which the experiences described by the items took place (frequency). Item scores ranged from 1 (never) to 7 (always) with higher total scores indicating a more positive alliance (Horvath and Greenberg, 1989). Data derived from the WAI-S were analysed to assess the client-therapist relationship (Watson et al., 2015). This WAI data were used in the current study to examine validity of the BATS.

#### **6.2.2.2 AESOPS**

AESOPS was a multicentre RCT, comparing the clinical and cost effectiveness of a stepped care intervention with a minimal intervention in the treatment of older hazardous alcohol users in primary care (Watson et al., 2013a). The stepped care intervention involved a 20-minute session of behaviour change counselling (BCC) (step 1), with referral to motivational enhancement therapy (MET; Miller et al., 1992) (step 2) and local specialist services (step 3) where necessary. BCC was a manual guided brief intervention using the technique of motivational interviewing (MI; Miller and Rollnick, 2013) to address clients' motivation to change drinking. The minimal intervention was 5-minutes of brief advice (BA), involving feedback of the results of the Alcohol Use Disorders Test (screening questionnaire), and discussion of the health consequences of continued hazardous alcohol use. BCC (step 1, stepped care) and BA (minimal) sessions were delivered by a practice/research nurse, or research practitioner, in primary care general practices in England and Wales.

In the trial, sessions were recorded and rated to assess treatment fidelity. Sessions were recorded onto tape and included audio data only. One hundred and sixty BCC and BA sessions were rated as part of the trial, 30% of the total number of sessions ( $n=529$ )<sup>75</sup>. Sessions were rated using the AESOPS PRS<sup>76</sup> (Tober and Crosby, 2011). The 18-item fidelity measure reflected the treatment components specified in the session protocols for BA, BCC, and MET (Watson et al., 2013a). Most items were scored on two 5-point scales, measuring the extent to which therapists carried out item behaviours (frequency), and how well therapists performed them (quality); item scores ranged from: 0 (not at all/very

---

<sup>75</sup> MET (step 2, stepped care) sessions were not rated, there were too few recordings to enable meaningful results to be obtained (Watson et al., 2013a).

<sup>76</sup> AESOPS PRS = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale.

poor) to 4 (extensively/very well). Data derived from the AESOPS PRS were analysed to ensure treatment fidelity (Watson et al., 2013a). Process rating data from the AESOPS trial were used in the validation of the BATS.

### **6.2.2.3 UKATT**

UKATT was a multicentre RCT, comparing the effectiveness of social behaviour and network therapy (SBNT) and motivational enhancement therapy (MET) for help seeking dependent drinkers (UKATT Research Team, 2005). SBNT was an integrative approach, combining cognitive and behavioural strategies, for problem drinkers (Copello et al., 2002). The intervention helped clients to build networks of people supportive of positive change in their drinking and associated behaviours (UKATT Research Team, 2001). MET was a manual-based adaptation of MI (Miller and Rollnick, 2013), and reinforced clients' motivation for and commitment to change their drinking (Miller et al., 1992). SBNT comprised eight 50-minute sessions, and MET three 50-minute sessions. Sessions were delivered by therapists working at treatment services around Birmingham, Cardiff, and Leeds.

In the trial, sessions were recorded and rated to assess treatment fidelity. Sessions were recorded onto DVD with visual recordings of therapists. Four hundred and fifty-two sessions were rated as part of the trial, 27% of the total number of sessions ( $n=1664$ ). Sessions were rated using the UKATT PRS<sup>77</sup> (Middleton et al., 2001). The 27-item scale covered the specific therapeutic techniques described in the treatment manuals (Tober et al., 2008). Most items were scored on two 5-point scales, measuring how often therapists carried out item behaviours (frequency), and how well therapists delivered them (quality); item scores ranged from: 0 (not at all/not at all well) to 4 (extensively/very well). Data derived from the UKATT PRS were used in the trial to ensure treatment fidelity (Tober et al., 2008). This process rating data were used in the current study to test the psychometric properties of the BATS.

In UKATT, the 12-item short form of the WAI (WAI-S; Tracey and Kokotovic, 1989) was used to examine changes in therapeutic alliance. The WAI-S was completed by both client and therapist at the end of the first and last therapy sessions (UKATT Research Team, 2001). Data derived from the WAI-S enabled examination of the treatment processes (UKATT

---

<sup>77</sup> UKATT PRS = UKATT (United Kingdom Alcohol Treatment Trial) Process Rating Scale.



Research Team, 2001). WAI data from the UKATT trial were used in the validation of the BATS.

### **6.2.3 Routine practice data**

Routine practice (RP) data comprised new recordings of therapy sessions delivered by therapists working at two treatment services for alcohol and drug use problems. Sessions were recorded at a NHS specialist addictions service (SAS) in the North of England, and a non-NHS drug and alcohol service (DAS) in Wales. Therapies delivered by these services were based on the principles of MI (Miller and Rollnick, 2013) and SBNT (Copello et al., 2002). An overview of the recordings and the recruitment procedure at each service is provided below.

#### **6.2.3.1 NHS specialist addictions service (SAS)**

##### *6.2.3.1.1 Recordings collected at the SAS*

Sixteen routine practice therapy sessions were recorded onto DVD, with visual recordings of therapists, between March 2015 and June 2015. Recordings involved seven therapists (four female, and three male), 16 clients (seven female, and nine male), and one friend or family member (FFM) (female) who attended a client's session.

Therapists were: five nurses, one psychologist, and one addictions therapist. The mean number of years therapists had worked in the addiction field was 7.6 years (standard deviation (SD): 5.9 years). All therapists were educated to degree level, with four therapists having gained postgraduate qualifications in a health-related discipline; for example, a non-medical prescribing post-graduate certificate. Five therapists listed a range of additional therapy specific courses they had attended; for example, cognitive behavioural skills.

Clients had a mean age of 39.8 years (SD: 4.8 years)<sup>78</sup>. Alcohol was the main problem substance for eight clients. Other primary problem substances included: buprenorphine ( $n=2$ ), cannabis ( $n=2$ ), crack cocaine ( $n=1$ ), and opiates ( $n=1$ )<sup>79</sup>. Eight of the clients had previously accessed treatment for substance use problems<sup>80</sup>.

---

<sup>78</sup> Missing data for 4 clients (age).

<sup>79</sup> Missing data for 2 clients (primary problem substance).

<sup>80</sup> Missing data for 4 clients (access to previous treatment).

#### *6.2.3.1.2 Recruitment procedure at the SAS*

All therapists working at the SAS, where treatment sessions were routinely recorded for supervision purposes, were potential participants. The researcher (HC) approached therapists to: explain the study, distribute copies of the participant information sheet (PIS) (Appendix E), and answer questions. Therapists wishing to take part completed a therapist consent form (Appendix E), and a short questionnaire asking about therapists' level of education, years worked in the addictions field, and any relevant training (Appendix E).

Therapists recording their treatment sessions for the research project were asked to identify clients currently on their caseload who could be approached to take part in the study. The decision to approach clients was informed by the study's eligibility criteria, and therapists' clinical judgement on the appropriateness of including individual clients. Clients were eligible for inclusion if they were willing and able to provide informed consent, were aged 18 years and over, and were accessing treatment for alcohol or drug use problems at the SAS.

SAS clients were routinely booked into clinic sessions at least a week before their appointment. An information pack was sent, by post, to those who were identified as eligible for inclusion and who were attending therapy sessions the following week. The pack included an introductory letter from the client's therapist (Appendix E), and a copy of the client PIS (Appendix E). Clients had at least two days to consider the PIS, ask questions, and talk to friends and family about their potential involvement.

At the treatment sessions, therapists explained the study, took consent (Appendix E), and recorded the sessions using DVD recorders. In the event of a FFM attending, therapists sought informed consent from both the client and the FFM; the PIS and consent forms used for FFMs were similar to those given to the client (Appendix E). After the session, therapists completed the client questionnaire, detailing the clients' age, gender, primary problem substance, and treatment history (Appendix E). Information on the recordings was entered into an Excel 2013 spreadsheet.

#### **6.2.3.2 Non-NHS drug and alcohol service (DAS)**

##### *6.2.3.2.1 Recordings collected at the DAS*

Nine routine practice therapy sessions were recorded onto DVD, with visual recording of

therapists, between December 2015 and June 2016. Recordings involved five therapists (two female, and three male), and nine clients (one female, and eight male).

Therapists included: three addiction therapists, one alcohol treatment worker, and one substance use keyworker. The mean number of years therapists' had worked in the addiction field was 4.5 years (SD: 3.3 years). Three therapists were educated to degree level, with two therapists having postgraduate qualifications in a health-related discipline (e.g., a counselling skills postgraduate certificate). All therapists had attended therapy specific training, including MI and SBNT.

Clients had a mean age of 49.9 years (SD: 8.3 years). Alcohol was the main problem substance for seven clients. Other primary problem substances included: benzodiazepines ( $n=1$ ), and opiates ( $n=1$ ). Three of the clients had previously accessed treatment for substance use problems.

#### *6.2.3.2.2 Recruitment procedure at the DAS*

DAS recruitment and consent processes followed similar procedures to the SAS. There were two notable differences. First, the researcher liaised with the Engagement Team Leader (ETL) about the study, rather than approaching therapists directly. The ETL approached therapists to: explain the study, distribute copies of the PIS, and answer questions. Second, clients were not sent an information pack prior to attending their appointment. Clients identified as eligible for inclusion were informed about the study at the start of their treatment session; copies of the PIS were distributed at this point. There were two reasons for not distributing the packs: i) SAS therapists reported that the packs had caused additional anxiety to clients and had impacted on treatment engagement, and ii) the DAS did not have administrative support. Discussing the PIS with therapists enabled clients to make an informed decision on whether to take part.

#### **6.2.4 Process rating data derived from the BATS**

So far, this chapter has described the trial and routine practice data used in the validation of the BATS. To conduct the reliability and validity analyses, the study also collected process rating data using the BATS. The next two sections describe how this process rating data was collected.

#### **6.2.4.1 Selecting recordings for process rating**

Trial and routine practice data comprised recordings of therapy sessions from three RCTs, and two treatment services for alcohol and drug use problems. Eighty of these recordings were randomly selected for independent process rating (rated by HC) (**Figure 13**). The BATS was designed to be transtheoretical, applicable to the range of widely used therapies in addiction. In routine practice, therapists use a range of psychological therapies, from brief advice to intensive specialist treatment, to address alcohol and drug use problems (Raistrick et al., 2006). To provide support for the utility of BATS in routine practice, it was important that the selected recordings were of different lengths and therapy types. The sample was, therefore, stratified by data source (AESOPS, ADAPTA, UKATT, and RP), and therapy (BA, BCC, AF, HL, MET, SBNT). Recordings from routine practice were not stratified by service; both treatment services (SAS and DAS) offered similar therapies. This sampling strategy also ensured that the selected recordings targeted clients' at different stages of therapy. Twenty of the 80 recordings were selected for double rating (rated by HC and GT), 25% of the total number of sessions (**Figure 13**).

To select the recordings for independent rating, information on the trial and routine practice recordings were compiled in an Excel 2013 spreadsheet, including: the data source, therapy/service, session number<sup>81</sup>, and session identification (ID) number<sup>82</sup>. Once entered into Excel, the recordings were randomly ordered using the 'RAND' function; the first 10 recordings for each trial intervention, and the first 20 routine practice recordings were selected. Replacement sampling was used for six missing trial recordings<sup>83</sup>. To select the recordings for double rating, data compiled in Excel was exported to IBM SPSS (Statistical Package for Social Sciences) Statistics, version 22 (IBM Corp., 2013); the recordings were randomly selected using the 'select cases' function.

Sample sizes were based on: i) previous work exploring psychometric properties within the addiction field (e.g., Lane et al., 2005; Torrey, 2012; Watson et al., 2015), and ii) what was feasible to do within the constraints of the project. Emphasis was given to the number of sessions required for testing inter-rater reliability; the most commonly reported test for

---

<sup>81</sup> For routine practice recordings, the session number was always '1'.

<sup>82</sup> For the trial recordings, the session ID matched the ID in the trial datasets.

<sup>83</sup> Missing recordings included: 1 BA session (AESOPS), 1 SBNT session (UKATT), and 4 MET sessions (UKATT).

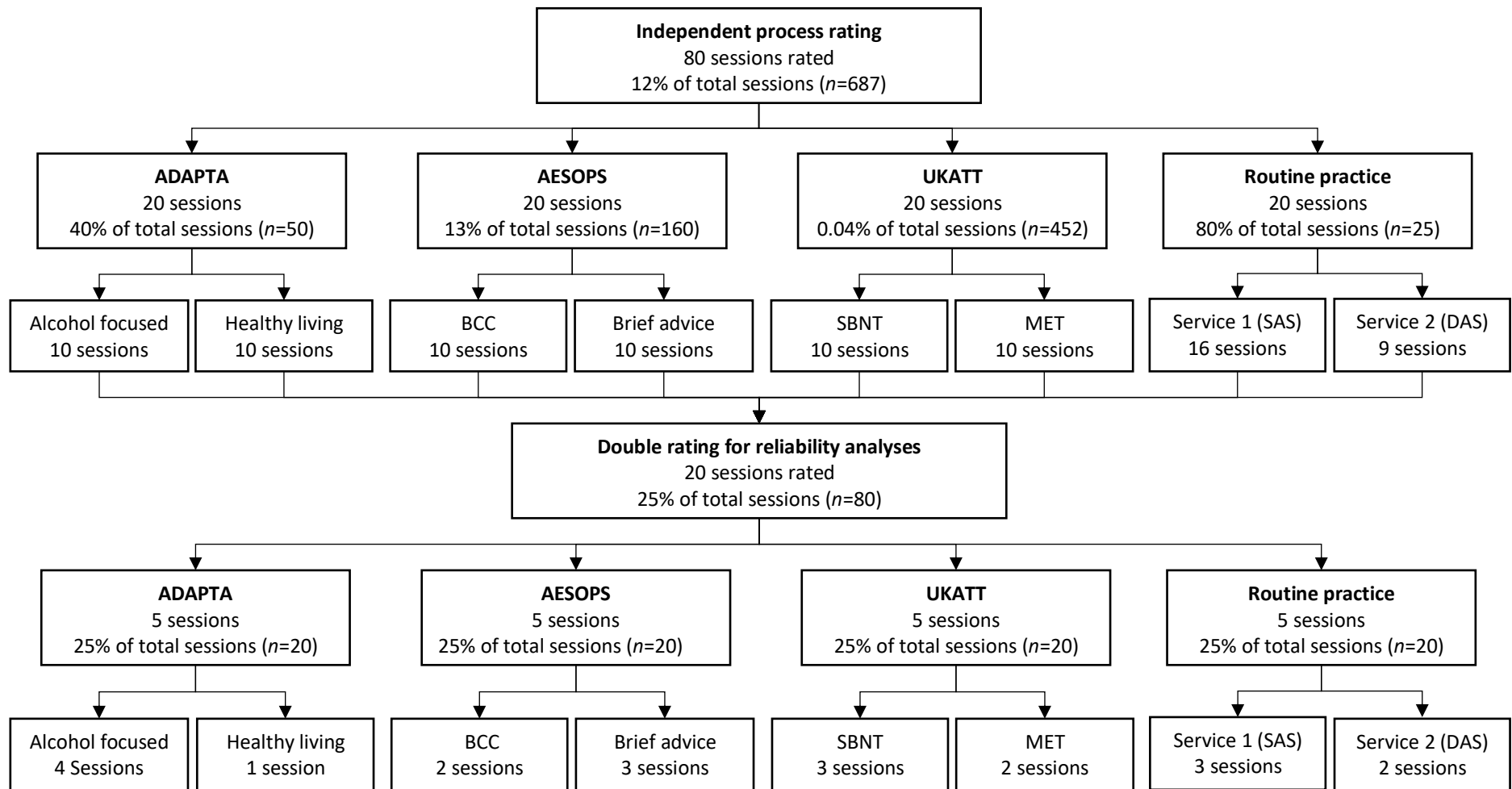


Figure 13: Flow diagram of the selection of recordings for independent and double rating

examining the psychometric properties of the measures identified in Study 1.

#### **6.2.4.2 The process of rating**

The newly developed BATS was used to rate the selected recordings. Raters listened to each recording, and scored the extent to which therapists demonstrated each of the item behaviours. Brief notes were made during the session to support item scoring; items were scored at the end of the session. Note taking was particularly useful for longer sessions and double rating. Sessions selected for double rating were rated independently by HC and GT. The rated tapes were then discussed with reference to the item definitions for the purposes of calibration (Watson et al., 2013a). Sessions were timed to record duration. Sessions were rated by HC from January 2017 to January 2018. Double rating took place at regular intervals throughout this time period. Session scores and duration were entered into IBM SPSS Statistics, version 22 (IBM Corp., 2013) (BATS dataset).

#### **6.2.5 Data analysis**

The development of the BATS involved the investigation of its validity and reliability; BATS was subjected to convergent validity and inter-rater reliability analyses. Duration of the rated sessions and the item scores were explored. Analyses were conducted in IBM SPSS Statistics, version 22 (IBM Corp., 2013), and Stata, version 13 (StataCorp, 2013b).

##### **6.2.5.1 Sessions rated using the BATS**

Duration of the session recordings was examined ( $n=80$ ). Descriptive statistics (mean, SD, and minimum and maximum duration) were calculated for each therapy type. A weighted least squares (WLS) linear regression model was used to compare session duration across the different therapy groups. Linear regression models can specify the differences between group means while avoiding the problem of multiple testing (Field, 2017). The dependent variable was session length, and the independent variable therapy type. Therapy type was nominal with seven categories: AF, HL, BCC, BA, SBNT, MET, and RP. The categories were, therefore, recoded to create six dummy variables (**Table 29**), which replaced 'therapy type' in the regression model. The sample size was considered appropriate; there were 10 or more cases in each independent (dummy) variable (Field, 2017).

Computing a linear regression enabled the assumptions of the ordinary least squares (OLS) method to be checked. The residuals were normally distributed; however, the variance showed heteroscedasticity (Appendix E). Given the non-constant variance, the analysis was

rerun using a WLS linear regression (Field, 2017). The weights were based on an auxiliary regression of the absolute residuals (dependent variable) versus the dummy variables (independent variables). The fitted (or predicted) values from the auxiliary regression were used to define the weights as 1 over the squared fitted values. The weights were incorporated into the linear regression model using the ‘WLS Weight’ option in SPSS (IBM Corp., 2013).

**Table 29: Coding design for the seven-category ‘therapy type’ variable**

Therapy Type	Dummy Variables					
	AF	HL	BCC	BA	SBNT	MET
AF	1	0	0	0	0	0
HL	0	1	0	0	0	0
BCC	0	0	1	0	0	0
BA	0	0	0	1	0	0
SBNT	0	0	0	0	1	0
MET	0	0	0	0	0	1
RP*	0	0	0	0	0	0

\* Reference category.

AF = Alcohol focused intervention (ADAPTA); BA = Brief advice (AESOPS); BCC = Behaviour change counselling (AESOPS); HL = Healthy living intervention (ADAPTA); MET = Motivational enhancement therapy (UKATT); RP = Routine practice therapy sessions; SBNT = Social behaviour and network therapy (UKATT).

#### **6.2.5.2 Item scores on the BATS**

Scores for items on the BATS were summarised by therapy type (median, interquartile range (IQR), range). Similar to the procedure described previously, a WLS linear regression model was used to compare total scores on the BATS across the therapies. The dependent variable was the BATS total scores, and the independent variables were the newly created dummy variables (**Table 29**). The dummy variables related to the six trial therapies (AF, HL, BCC, BA, SBNT, and MET) with RP as the reference category. A WLS procedure was used because the residuals followed a normal distribution with non-constant variance.

#### **6.2.5.3 Convergent validity**

Convergent validity was examined by correlating scores on the BATS with five criterion measures: client and therapist versions of the 12-item short form of the WAI (WAI-S; Tracey and Kokotovic, 1989), the ADAPTA PRS (Tober and Crosby, 2014), the AESOPS PRS, (Tober and Crosby, 2011) and the UKATT PRS (Middleton et al., 2001).

##### **6.2.5.3.1 Relationship to the Working Alliance Inventory**

WAI data from ADAPTA and UKATT were used to examine convergent validity of the BATS. In both trials, data were collected at two time points using client and therapist forms of the WAI-S. WAI data from one time point were merged with the BATS dataset using session identifiers. The time point closest to corresponding sessions was selected (Appendix E); for example, the BATS dataset included scores for a fourth AF session from ADAPTA, corresponding WAI data collected at the end of session 3 (rather than session 2) was merged with the BATS dataset. There were two reasons for using WAI data for both client and therapist: i) studies indicate that client evaluations of the therapeutic alliance are better predictors of treatment outcomes than therapist-rated alliance (e.g., Mee-Lee et al., 2010; Cook et al., 2015), and ii) the BATS evaluated therapist delivery of treatments, client behaviours were not assessed.

Convergent validity was examined in two ways. First, total scores on the BATS were compared with total scores on the WAI-S for client and therapist. Second, item scores for 'collaboration', 'empathy' and 'complex reflections' on the BATS were compared with total scores on the WAI-S for client and therapist. It was hypothesised that the support items would correlate more highly with the WAI-S than the total scores on the BATS; the alliance concerns the collaborative relationship between client and therapist (Horvath and Greenberg, 1989). A weaker relationship was expected between total scores on the BATS and the WAI-S because the BATS is not a measure of the therapeutic alliance.

Scores were compared using the non-parametric correlation coefficient, Kendall's tau-b, as items on the BATS were measured at the ordinal level. Kendall's tau-b is a measure of "*the difference between the probability that the observed data are in the same order versus the probability that the observed data are not in the same order*" (Bowers, 2014, p.259). This correlation coefficient was considered advantageous to Spearman's rho, a more popular non-parametric rank correlation, for analysing small samples with tied ranks (Arndt et al., 1999; Field, 2017). Appendix E provides an example calculation.

Bootstrap confidence intervals (CIs) were computed. Bootstrapping involves repeated sampling from the original sample of data. The test statistic, in this case Kendall's tau-b, is calculated for each bootstrap sample and the results amalgamated (Bowers, 2014). CIs are estimated from the sampling distribution of the test statistic for the original sample (Wright et al., 2011). The bias-corrected accelerated (BCa) CI was selected, a more accurate method for calculating bootstrap CIs than the alternative percentile interval



(Field, 2017).

The correlation coefficients and bootstrap CIs were computed using SPSS (IBM Corp., 2013). The strength of the relationships was evaluated using the commonly used effect sizes outlined in Field (2017) (**Table 30**).

**Table 30: Assessing associations with Kendall's tau-b**

$\tau_b$	Strength of relationship
$<\pm 0.30$	Weak
$\pm 0.30$ to $\pm 0.49$	Moderate
$\pm 0.50$ to $\pm 1.00$	Strong

$\tau_b$  = Kendall's tau-b; < = Less than.

#### 6.2.5.3.2 Relationship to the fidelity measures

Process rating data from ADAPTA, AESOPS, and UKATT were used to examine convergent validity of the BATS. In all three trials, the fidelity measures (ADAPTA PRS, AESOPS PRS, and UKATT PRS) scored item frequency and quality: the extent to which therapists carried out item behaviours, and the quality with which they were performed. Data derived from the measures were merged with the BATS dataset using session identifiers. Frequency and quality scores were included for two reasons. First, items on the BATS measure the extent to which behaviours were carried out. Second, the relationship between adherence and competence is unresolved (Perepletchikova et al., 2007); the BATS may have potential to impact on therapist competence, and client treatment outcomes (Campos-Melady et al., 2017).

Convergent validity was examined in two ways: i) total scores on the BATS were compared with total scores on the ADAPTA PRS, AESOPS PRS, and UKATT PRS for item frequency and quality, and ii) total scores on the BATS were compared with scores for selected items on each of the three measures. Selected items evaluated therapist activities that were considered similar to those described in the BATS (Appendix E); for example, all three measures included an item on therapist empathy, these items were selected as this behaviour was also evaluated in the BATS. It was hypothesised that the BATS total scores would correlate more highly with scores for the selected items than total scores on the process rating measures.

The ADAPTA dataset contained six missing scores for item 'plan behaviour' frequency and quality. Regression imputation was conducted to predict the missing values from the regression equation (Bland, 2017). However, the models (for frequency and quality) failed to converge. Given the small sample sizes, case deletion was not considered appropriate. Therefore, item scores were excluded from the analyses; total scores and scores for selected items on the ADAPTA PRS excluded item scores for 'plan behaviour'.

Scores were compared using the correlation coefficient, Kendall's tau-b. Three variables were non-normally distributed: i) ADAPTA frequency scores for selected items, ii) AESOPS quality total scores, and iii) UKATT quality scores for selected items. Visual checks of normality were supported by the Shapiro-Wilk test (Appendix E). BCa 95% bootstrap CIs for the correlation coefficients were computed using SPSS (IBM Corp., 2013). The coefficients were evaluated using the effect sizes outlined in Field (2017) (**Table 30**).

#### **6.2.5.4 Inter-rater reliability**

A sample of the recordings was rated by two raters (HC and GT) ( $n=20$ ). Inter-rater reliability (IRR) of the item scores was examined using two tests: the weighted kappa, and the intraclass correlation coefficient. The weighted kappa is a measure of agreement for ordinal variables (Mandrekar, 2011) – items on the BATS were measured at the ordinal level. By contrast, the intraclass correlation coefficient (ICC) is a test of agreement for continuous variables (Liu et al., 2016). ICCs were commonly used in the validation of the fidelity measures identified from the literature review in Study 1 (e.g., Carroll et al., 2000; Martino et al., 2008; Watson et al., 2013a). ICCs were therefore used, in addition to the weighted kappa, to aid comparison with the previous literature.

##### **6.2.5.4.1 Weighted kappa**

Weighted kappa ( $k_w$ ) adjusts for the degree of disagreement between ordinal categories (Warrens, 2012). For example, the disagreement between 'not at all' and 'very little' is not as great as between 'not at all' and 'extensively'. Adjustments are made by allocating weights which reflect differences in the magnitude of scoring discrepancies between the raters (Altman, 1999; Mandrekar, 2011). There are two commonly used weights: linear weights, and quadratic weights (Brenner and Kliebsch, 1996). Quadratic weights were used for two reasons. First, the levels of disagreement were not considered equal, larger relative penalties were attached to larger disagreements (Bland, 2017). Second, the weighted kappa can, in some cases, be interpreted as an ICC when using quadratic weights (Warrens,

2012; Hallgren, 2012); the interchangeability of the two statistics has implications for measuring and reporting IRR for ordinal data. Quadratic weights were calculated as weights for the agreement, rather than the disagreement (Bland, 2017) (Table 31). Appendix E provides an example calculation.

**Table 31: Ratings weighted with quadratic weights**

Rater 1	Rater 2				
	Not at all	Very little	Somewhat	A good deal	Extensively
Not at all	<b>1.00</b>	0.9375	0.75	0.4375	0.00
Very little	0.9375	<b>1.00</b>	0.9375	0.75	0.4375
Somewhat	0.75	0.9375	<b>1.00</b>	0.9375	0.75
A good deal	0.4375	0.75	0.9375	<b>1.00</b>	0.9375
Extensively	0.00	0.4375	0.75	0.9375	<b>1.00</b>

The weighted kappa and BCa 95% bootstrap CIs were calculated using Stata, version 13 (StataCorp, 2013b). The reliability coefficients were assessed using guidelines outlined in Bowers et al. (2014) (Table 32).

**Table 32: Assessing agreement with weighted kappa**

$k_w$	Strength of agreement
<0.20	Poor
0.20 to 0.40	Fair
0.41 to 0.60	Moderate
0.61 to 0.80	Good
0.81 to 1.00	Very good

$k_w$  = weight kappa; < = less than.

#### 6.2.5.4.2 Intraclass correlation coefficient (ICC)

ICCs are calculated as the ratio of variability between therapists to the total variability, where the total variability includes both therapist and error variability (Streiner et al., 2015). ICCs using a two-way mixed-effects model (3,1) (Shrout and Fleiss, 1979) were computed. This model was chosen for four reasons. First, recordings were rated by the same two raters; there was an effect of the therapist (from the recording) and an effect of the raters (two-way model) (Hallgren, 2012). Second, the raters were not randomly selected; the therapists were a random effect (therapy sessions were randomly selected), and the raters a fixed effect (mixed-effects). Third, the reliability of the scores for a single rater was examined; the BATS was designed to be completed by individual therapists, ICCs

based on the mean of multiple raters would be misleadingly high (Streiner et al., 2015). Lastly, reliability was defined in terms of consistency, rather than absolute agreement; the chosen model corresponds to assessing consistency when the raters are the only raters of interest in the study, a fixed effect (Streiner et al., 2015; Trevethan, 2017). Appendix E provides an example calculation.

ICC estimates and their 95% CIs were computed using SPSS (IBM Corp., 2013), and were interpreted using guidelines provided by Cicchetti (1994) (**Table 33**).

**Table 33: Assessing agreement with ICC**

ICC	Strength of agreement
<0.40	Poor
0.40 to 0.59	Fair
0.60 to 0.74	Good
0.75 to 1.00	Excellent

ICC = intraclass correlation coefficient; < = less than.

## 6.3 Results

### 6.3.1 Sessions rated using the BATS

Descriptive statistics of the duration of sessions for each therapy type are presented in **Table 34**. The mean duration ranged from 6 minutes (BA) to 55 minutes (SBNT). A WLS linear regression showed a statistically significant difference in the session length across the therapies,  $F(6,73)=19.00$ ,  $p<0.001$ .

**Table 35** compares the duration of sessions between RP and each trial therapy. There were statistically significant differences for the AESOPS and UKATT therapies (BCC, BA, SBNT, and MET). The largest differences were found between RP and BA, and between RP and SBNT; the duration of RP was significantly longer than the duration of BA, and significantly shorter than SBNT. The smallest differences were found for RP and the ADAPTA therapies (AF and HL); these differences were not statistically significant. These descriptive statistics highlight the range of therapies used in the validation of the BATS.

**Table 34: Summary of the length of the sessions**

Therapy (Planned Duration)	N	Duration of the Sessions (hours: minutes: seconds)			
		Mean	SD	Minimum	Maximum
AF (30-45 minutes)	10	0:43:51	0:16:39	0:22:27	1:14:29
HL (30-45 minutes)	10	0:40:01	0:10:44	0:26:27	0:53:29
BCC (20 minutes)	10	0:20:28	0:03:29	0:15:56	0:26:28
BA (5 minutes)	10	0:06:17	0:01:46	0:04:21	0:09:24
SBNT (50 minutes)	10	0:54:35	0:08:07	0:38:00	1:03:52
MET (50 minutes)	10	0:50:49	0:19:17	0:18:52	1:27:47
RP (unknown)	20	0:32:53	0:13:39	0:14:14	0:54:05
<b>Total</b>	<b>80</b>	<b>0:35:13</b>	<b>0:19:09</b>	<b>0:04:21</b>	<b>1:27:47</b>

N = Number of recordings; SD = Standard deviation; AF = Alcohol focused intervention (ADAPTA); BA = Brief advice (AESOPS); BCC = Behaviour change counselling (AESOPS); HL = Healthy living intervention (ADAPTA); MET = Motivational enhancement therapy (UKATT); RP = Routine practice therapy sessions; SBNT = Social behaviour and network therapy (UKATT).

**Table 35: Differences in duration between routine practice and trial therapies**

Comparison	Mean Duration*		Mean Difference ( <i>b</i> )	95% CI	<i>p</i> -value
	RP	Comparator			
RP vs. AF	0:32:53	0:43:51	657.6	-49.77 to 1361.67	0.067
RP vs. HL	0:32:53	0:40:01	428.1	-148.96 to 1005.16	0.144
RP vs. BCC	0:32:53	0:20:28	-745.2	-1157.76 to -332.64	0.001
RP vs. BA	0:32:53	0:06:17	-1596.1	-1992.49 to -1199.71	<0.001
RP vs. SBNT	0:32:53	0:54:35	1219.3	744.22 to 1694.38	<0.001
RP vs. MET	0:32:53	0:50:49	1075.5	280.76 to 1870.24	0.009

\* Mean length of the sessions (hours: minutes: seconds).

*b* = Regression coefficient; CI = Confidence interval; AF = Alcohol focused intervention (ADAPTA); BA = Brief advice (AESOPS); BCC = Behaviour change counselling (AESOPS); HL = Healthy living intervention (ADAPTA); MET = Motivational enhancement therapy (UKATT); RP = Routine practice therapy sessions; SBNT = Social behaviour and network therapy (UKATT).

### 6.3.2 Item scores on the BATS

A summary of the item scores on the BATS across the therapies are presented in **Table 36**. Item scores for RP and the ADAPTA therapies were relatively similar (AF and HL), as were the scores for BCC (AESOPS) and MET (UKATT). AF and SBNT had higher median scores for item 12 'sources of support' compared to the other therapies. BA had the lowest scores for the majority of items; fewer therapist behaviours were carried out in the 5 minutes of BA compared to the other therapies, which were longer in length. Item 9 'developing discrepancy' had low scores across all therapy types.

**Table 36: Summary of the item scores across therapies**

Item reference	Therapy							
	AF (n=10)	HL (n=10)	BA (n=10)	BCC (n=10)	MET (n=10)	SBNT (n=10)	RP (n=20)	Total (n=80)
	Median Scores* (IQR), Range	Median Scores* (IQR), Range	Median Scores* (IQR), Range	Median Scores* (IQR), Range	Median Scores* (IQR), Range	Median Scores* (IQR), Range	Median Scores* (IQR), Range	Median Scores* (IQR), Range
1. Problem focused	4 (0), 3-4	3.5 (2), 1-4	3.5 (1), 1-4	3 (2), 2-4	3 (2), 1-4	2 (2), 1-4	4 (1), 1-4	3 (1), 1-4
2. Collaboration	3.5 (1), 3-4	4 (1), 3-4	1 (1), 0-2	2 (2), 1-4	2 (3), 0-4	1 (2), 0-4	4 (1), 2-4	3 (2), 0-4
3. Empathy	3 (1), 2-4	4 (1), 3-4	2 (1), 1-3	2.5 (1), 2-4	2 (1), 1-4	2 (2), 1-4	4 (1), 2-4	3 (2), 0-4
4. Strengths and affirmation	1.5 (3), 0-4	1 (2), 0-4	0 (0), 0-1	1 (1), 0-2	1 (2), 0-3	1 (3), 0-3	2 (2), 0-4	1 (2), 0-4
5. Complex reflections	3 (2), 2-4	3 (1), 2-4	0 (1), 0-1	2 (1), 1-3	2 (2), 1-4	1 (1), 1-4	2 (3), 0-4	2 (2), 0-4
6. Homework assigned	2 (2), 0-3	1 (1), 1-4	0 (0), 0-0	0 (0), 0-0	0 (0), 0-1	0.5 (0), 0-1	1.5 (3), 0-4	0 (2), 0-4
7. Homework reviewed	1 (1), 0-3	0.5 (3), 0-4	0 (0), 0-0	0 (0), 0-0	0 (0), 0-3	0 (0), 0-1	1 (1), 0-4	0 (2), 0-4
8. Treatment goals	3.5 (1), 1-4	2 (2), 1-4	0 (1), 0-1	2.5 (2), 0-4	2 (2), 1-4	2 (2), 0-4	3 (2), 0-4	2 (2), 0-4
9. Developing discrepancy	0 (1), 0-2	0 (0), 0-1	0 (0), 0-0	0 (0), 0-1	1 (2), 0-2	0 (1), 0-1	0 (0), 0-3	0 (1), 0-3
10. Exploring pros and cons of change	1 (2), 0-2	1 (2), 0-3	0 (0), 0-1	1 (1), 0-2	1.5 (3), 0-4	0 (0), 0-1	1 (2), 0-4	1 (2), 0-4
11. Behaviour change planning	1 (2), 0-3	2 (2), 1-4	0 (0), 0-1	1 (1), 0-3	0.5 (2), 0-3	1.5 (2), 1-4	2 (2), 0-4	1 (3), 0-4
12. Sources of support	3 (0), 2-4	1 (1), 0-4	0 (0), 0-1	0 (0), 0-0	0.5 (1), 0-2	2 (3), 1-4	1 (1), 0-4	1 (2), 0-4

\*Scores made on a 5-point scale: 0=not at all, 1=a little, 2=somewhat, 3=a good deal, 4=extensively.

IQR = Interquartile range; AF = Alcohol focused intervention (ADAPTA); BA = Brief advice (AESOPS); BCC = Behaviour change counselling (AESOPS); HL = Healthy living intervention (ADAPTA); MET = Motivational enhancement therapy (UKATT); RP = Routine practice therapy sessions; SBNT = Social behaviour and network therapy (UKATT).

Total scores on the BATS for each therapy are summarised in Error! Not a valid bookmark self-reference.; the mean score ranged from 7.6 (BA) to 27 (AF). A WLS linear regression showed a statistically significant difference in total scores on the BATS across the therapies,  $F(6)=45.028$ ,  $p<0.001$ . There were statistically significant differences between RP and the AESOPS and UKATT therapies (BCC, BA, SBNT, and MET). The largest differences were found for BA and BCC; RP total scores were significantly higher than BA total scores and BCC total scores. The smallest differences were found for the ADAPTA therapies (AF and HL); these differences were not statistically significant.

**Table 37: Differences in total scores between routine practice and trial therapies**

Comparison	Mean Total Score (SD)		Mean Difference ( <i>b</i> )	95% CI	<i>p</i> -value
	RP	Comparator			
RP vs. AF	25.80 (7.30)	27.00 (4.40)	1.20	-3.27 to 5.67	0.594
RP vs. HL	25.80 (7.30)	25.30 (7.68)	-0.05	-6.33 to 5.33	0.865
RP vs. BCC	25.80 (7.30)	15.10 (4.84)	-10.70	-15.22 to -6.18	<0.001
RP vs. BA	25.80 (7.30)	7.60 (1.84)	-18.20	-21.65 to -14.75	<0.001
RP vs. SBNT	25.80 (7.30)	16.40 (6.48)	-9.40	-14.38 to -4.42	<0.001
RP vs. MET	25.80 (7.30)	17.20 (7.42)	-8.60	-14.34 to -2.86	0.004

SD = Standard deviation; *b* = Regression coefficient; CI = Confidence interval; AF = Alcohol focused intervention (ADAPTA); BA = Brief advice (AESOPS); BCC = Behaviour change counselling (AESOPS); HL = Healthy living intervention (ADAPTA); MET = Motivational enhancement therapy (UKATT); RP = Routine practice therapy sessions; SBNT = Social behaviour and network therapy (UKATT).

### 6.3.3 Convergent validity

#### 6.3.3.1 Relationship to the Working Alliance Inventory

WAI data from ADAPTA and UKATT were used to examine convergent validity of the BATS.

**Table 38** shows the number of cases in the BATS dataset with corresponding WAI data. The relationships between the BATS and the WAI-S for both client and therapist were examined using Kendall's tau-b (**Table 39**, and **Table 40**).

**Table 38: Number of cases in the BATS dataset with corresponding WAI data**

Data Source	Cases	Trial		
		ADAPTA	UKATT	Total
BATS	N	20	20	40
WAI-S Client	N	11	13	24
WAI-S Therapist	N	15	16	31

N = Number of cases.

**Table 39: Relationships between the BATS and the WAI-S client (n=24)**

	WAI Client Total Scores		
	$\tau_b$	<i>p</i> -value	BCa 95% Bootstrap CI
BATS Total Scores	0.40	0.007	0.09 to 0.67
BATS Item Scores for 'Collaboration'	0.50	0.002	0.19 to 0.67
BATS Item Scores for 'Empathy'	0.38	0.020	0.03 to 0.63
BATS Item Scores for 'Complex reflections'	0.29	0.071	-0.10 to 0.63

$\tau_b$  = Kendall's tau-b; BCa = Bias-corrected accelerated; CI = Confidence interval.

There were moderate, positive correlations between total scores on the BATS and total scores on the WAI-S for both client and therapist; the relationships were statistically significant. Item scores for 'collaboration' correlated strongly with total scores on the WAI-S client, and moderately with the WAI-S therapist. The positive correlations were statistically significant. Moderate, positive correlations were found between item scores for 'empathy' and total scores on the WAI-S for both client and therapist. The correlations were statistically significant. Item scores for 'complex reflections' correlated weakly with total scores on the WAI-S client, and moderately with WAI-S therapist. The positive correlations were statistically significant for the WAI-S therapist only ( $p=0.004$ ).

**Table 40: Relationships between the BATS and the WAI-S therapist (n=31)**

	WAI Therapist Total Scores		
	$\tau_b$	<i>p</i> -value	BCa 95% Bootstrap CI
BATS Total scores	0.33	0.011	0.09 to 0.53
BATS Item Scores for 'Collaboration'	0.36	0.011	0.09 to 0.59
BATS Item Scores for 'Empathy'	0.37	0.010	0.13 to 0.58
BATS Item Scores for 'Complex reflections'	0.41	0.004	0.16 to 0.61

$\tau_b$  = Kendall's tau-b; BCa = Bias-corrected accelerated; CI = Confidence interval.

The 95% bootstrap CIs supported the direction of the correlation coefficients, excepting the relationship between item scores for 'complex reflections' and total scores on the WAI-S client. For this exception, the 95% bootstrap CI supports the statistical non-significance of the coefficient (the CI contains '0), the relationship could be positive, negative, or none. The 95% bootstrap CIs were all relatively wide, indicating some uncertainty about the magnitude of the true (population) correlation coefficient (Bowers et al., 2014). For example, the 95% bootstrap CI for the strongest association of 0.50 was 0.19 to 0.67 (item scores for 'collaboration' and WAI-S client total scores). Because the CI represents the range of plausible values for the true coefficient, the effect size could be weak, moderate,



or strong.

### 6.3.3.2 Relationship to the process rating measures

Process rating data from ADAPTA, AESOPS, and UKATT were used to examine convergent validity of the BATS. The relationships between scores on the trial process rating measures and the BATS were examined using Kendall's tau-b ( $\tau_b$ ) (Table 41).

**Table 41: Relationships between the trial-specific fidelity measures and the BATS (n=20)**

PRS	Data	BATS Total Scores		
		$\tau_b$	<i>p</i> -value	BCa 95% Bootstrap CI
ADAPTA*	Total <i>f</i> scores	0.44	0.009	0.08 to 0.73
	Total <i>q</i> scores	0.44	0.010	0.04 to 0.73
	Summed <i>f</i> scores for selected items	0.42	0.012	0.06 to 0.73
	Summed <i>q</i> scores for selected items	0.39	0.020	0.04 to 0.69
AESOPS	Total <i>f</i> scores	0.50	0.003	0.15 to 0.75
	Total <i>q</i> scores	0.44	0.009	0.05 to 0.74
	Summed <i>f</i> scores for selected items	0.66	<i>p</i> <0.001	0.41 to 0.84
	Summed <i>q</i> scores for selected items	0.40	0.020	0.02 to 0.70
UKATT	Total <i>f</i> scores	0.24	0.309	-0.14 to 0.53
	Total <i>q</i> scores	0.19	0.419	-0.23 to 0.60
	Summed <i>f</i> scores for selected items	0.21	0.385	-0.24 to 0.54
	Summed <i>q</i> scores for selected items	0.17	0.475	-0.29 to 0.59

\*Scores for item 'plan behaviour change' were excluded from the analyses due to missing data. BCa = Bias-corrected accelerated; CI = Confidence interval; *f* = frequency; PRS = process rating scale; *q* = quality;  $\tau_b$  = Kendall's tau-b.

There were moderate correlations between total scores on the BATS and scores on the ADAPTA PRS. The positive correlations were statistically significant. Total scores on the BATS did not correlate more highly with scores for selected items than total scores on the ADAPTA PRS. Scores from the AESOPS PRS followed a similar pattern to ADAPTA. However, there were three differences. First, there were strong correlations between the BATS total scores and frequency scores on the AESOPS PRS (selected items:  $\tau_b=0.66$ , *p*<0.001; total scores:  $\tau_b=0.50$ , *p*=0.003). Second, the effect sizes for the quality scores were lower than the frequency scores. Third, the BATS total scores correlated more highly with frequency scores for selected items than frequency total scores on the AESOPS PRS.

There were weak correlations between total scores on the BATS and scores on the UKATT PRS. The positive associations were not statistically significant. The BATS total scores did

not correlate more highly with scores for the selected items than total scores on the UKATT PRS; a finding similar to ADAPTA. The effect sizes for the quality scores were lower than the frequency scores; a finding similar to AESOPS.

The 95% bootstrap CIs support the statistical significance of the correlation coefficients. The CIs were all relatively wide, there is uncertainty about the magnitude and direction of the true coefficient (Bowers et al., 2014). The small sample sizes ( $n=20$ ) may explain the lack of precision for the correlation coefficients (du Prel et al., 2009).

#### **6.3.4 Inter-rater reliability**

A sample of the recordings was rated by two raters (HC & GT) ( $n=20$ ). IRR of the items was first examined using the quadratic weighted kappa (**Table 42**). Reliability coefficients for the items ranged from 0.74 to 0.96, indicating good to very good agreement between the two raters' scores. The BCa 95% bootstrap CIs indicated some uncertainty about the magnitude of the effect, particularly for items 8, 10, and 11. It was, therefore, concluded that the IRR of the item scores was moderate to very good.

IRR of the items was also examined using the ICC two-way mixed-effects model (3.1) (Shrout and Fleiss, 1979). A similar pattern to the weighted kappa was found; ICCs for the items ranged from 0.74 to 0.96, indicating good to excellent levels of agreement (**Table 42**). The 95% CIs indicated some uncertainty about the magnitude of effect, particularly for items, particularly for items 8 and 11. It was, therefore, concluded that the IRR of the items was fair to excellent.

For the majority of items, the weighted kappa provided a more conservative estimate than the ICC, although the strength of agreement was similar for seven of the eight items. The weighted kappa was equivalent to the ICC for the remaining items.

## **6.4 Discussion**

### **6.4.1 Study overview**

The development of the BATS involved the investigation of its psychometric properties. Data used in the investigation included: secondary analysis of trial data, routine practice data, and process rating data derived from the BATS. Three RCTs were used in the secondary analysis of trial data: ADAPTA (Watson et al., 2015), AESOPS

**Table 42: Weighted kappa and intraclass correlation coefficients for items on the BATS**

Item Reference	Median Scores* (IQR)		$K_w$ (BCa 95% Bootstrap CI)	$ICC$ (95% CI)	$K_w$ vs. $ICC$
	Rater 1	Rater 2			
1. Problem focused	4 (3,4)	4 (3,4)	0.86 (0.63 to 0.95)	0.87 (0.69 to 0.94)	Lower $K_w$ , equivalent strength
2. Collaboration	3 (2.25,4)	4 (2,4)	0.93 (0.86 to 0.99)	0.93 (0.84 to 0.97)	Equivalent
3. Empathy	3 (2,4)	3 (1.25,4)	0.89 (0.80 to 0.95)	0.93 (0.82 to 0.97)	Lower $K_w$ , equivalent strength
4. Strengths and affirmation	2 (0.25,3)	3 (0.25,3.75)	0.96 (0.89 to 0.99)	0.96 (0.90 to 0.98)	Equivalent
5. Complex reflections	2 (1,3)	1.5 (0,3)	0.78 <sup>a</sup> (0.55 to 0.93)	0.85 (0.65 to 0.94)	Lower $K_w$ , different strength
6. Homework assigned	0 (0,1.75)	0.5 (0,2)	0.90 (0.64 to 1.00)	0.92 (0.80 to 0.97)	Lower $K_w$ , equivalent strength
7. Homework reviewed	0 (0,1)	0 (0,1)	0.87 (0.70 to 0.94)	0.88 (0.73 to 0.95)	Lower $K_w$ , equivalent strength
8. Treatment goals	3 (1,4)	2.5 (1,3.75)	0.73 <sup>a</sup> (0.42 to 0.93)	0.74 <sup>b</sup> (0.45 to 0.89)	Lower $K_w$ , equivalent strength
9. Developing discrepancy	0 (0,1)	0 (0,1)	0.83 (0.71 to 0.92)	0.87 (0.70 to 0.95)	Lower $K_w$ , equivalent strength
10. Exploring pros and cons of change	1 (0,2)	1 (0,2)	0.82 (0.53 to 0.97)	0.82 (0.60 to 0.93)	Equivalent
11. Behaviour change planning	1 (0.25,3)	2 (0.25,3)	0.80 (0.60 to 0.94)	0.81 (0.58 to 0.92)	Lower $K_w$ , equivalent strength
12. Sources of support	1 (0.25,2.75)	1.5 (0,2.75)	0.91 (0.76 to 0.97)	0.91 (0.79 to 0.96)	Equivalent

\*Scores made on a 5-point Likert scale: 0=not at all, 1=a little, 2=somewhat, 3=a good deal, 4=extensively.

<sup>a</sup> Good agreement for items 5 and 8; very good agreement for the remaining items.

<sup>b</sup> Good agreement for item 8; excellent agreement for the remaining items.

IQR = Interquartile range (25<sup>th</sup>, 75<sup>th</sup> percentiles);  $K_w$  = Kendall's tau-b; BCa = Bias-corrected accelerated; CI = Confidence interval;  $ICC$  = intraclass correlation coefficient.

(Watson et al., 2013a), and UKATT (UKATT Research Team, 2005). Trial data comprised: recordings of therapy sessions<sup>84</sup>, process rating data (collected using trial-specific fidelity measures), and WAI outcome data (ADAPTA and UKATT only). Routine practice data comprised new recordings of therapy sessions delivered by therapists working at two treatment services for alcohol and drug use problems: a NHS specialist addictions service (SAS) in the North of England, and a non-NHS drug and alcohol service (DAS) in Wales. Process rating data derived from the BATS comprised ratings (collected using the BATS) of therapy sessions selected from the routine practice and trial data.

Eighty recordings of the trial and RP therapy sessions were randomly selected for independent process rating. The selected sessions were rated by HC using the BATS. WLS linear regression models were used to examine the length of sessions and the item scores across the therapies. Convergent validity of the BATS was examined by merging the WAI and process rating datasets, and correlating scores on the BATS against scores on four criterion measures: the 12-item short form of the WAI (WAI-S; Tracey and Kokotovic, 1989), the ADAPTA PRS (Tober and Crosby, 2014), the AESOPS PRS, (Tober and Crosby, 2011) and the UKATT PRS (Middleton et al., 2001). Scores were compared using the non-parametric correlation coefficient, Kendall's tau-b.

Twenty of the recordings were selected for double rating; sessions were independently rated by HC and GT. Inter-rater reliability, the degree of consistency or agreement among the raters was examined using two tests: the weighted kappa, and the intraclass correlation coefficient (ICC). ICCs were used, in addition to the weighted kappa, to aid comparison with the previous literature.

## **6.4.2 Main findings**

### **6.4.2.1 Sessions rated using the BATS**

The eighty selected recordings covered seven different types of therapy, including sessions delivered in routine practice settings. The seven therapy types were: brief advice (BA), behaviour change counselling (BCC), an alcohol focused intervention (AF), a healthy living intervention (HL), motivational enhancement therapy (MET), social behaviour and network therapy (SBNT), and routine practice therapy sessions (RP). The mean duration of each

---

<sup>84</sup> ADAPTA: Video recordings of AF and HL sessions; AESOPS: Audio recordings of BCC and BA sessions; UKATT: Video recordings of SBNT and MET sessions.

therapy type ranged from 6 minutes (BA) to 55 minutes (SBNT). A WLS linear regression showed a statistically significant difference in the session length across the therapies,  $F(6,73)=19.00, p<0.001$ .

#### **6.4.2.2 Item scores on the BATS**

A WLS linear regression showed a significantly significant difference in the BATS total scores across the therapies. The mean total scores ranged from 7.6 (BA) to 27 (AF). Compared to the RP therapy sessions, the largest differences were found for the AESOPS therapies; RP total scores were significantly higher than total scores for both BA and BCC. The smallest differences were found for the ADAPTA therapies; these differences were not statistically significant. Total scores on the BATS reflected the item scores. For example, item scores for RP were similar to the ADAPTA therapies. BA had the lowest scores for the majority of items; therapists delivering 5-minutes of BA carried out fewer item behaviours than in the other therapies. Of note, item 9 'developing discrepancy' was scored consistently low across all therapies. There are two possible explanations: i) therapists were not carrying out the behaviour, and ii) therapists were developing a discrepancy between the clients' values/beliefs and substance use behaviours but used processes not evaluated in the BATS (Moyers et al., 2005). That being said, the analyses show that the BATS is able to distinguish between therapies with different content.

#### **6.4.2.3 Convergent validity**

##### *6.4.2.3.1 Relationship to the Working Alliance Inventory*

There were moderate, positive correlations between total scores on the BATS and total scores on the WAI-S for both client and therapist. The correlations were statistically significant. The findings provide evidence of convergent validity; the BATS and the WAI-S described similar behaviours, focusing on the collaborative relationship between client and therapist.

Of the support items (collaboration, empathy, complex reflections), only item scores for 'collaboration' showed a stronger effect than the BATS total scores against the WAI-S client. For the WAI-S therapist, scores for all three items were more highly correlated than the BATS total scores, although the effects were marginal. The mixed findings reflect the literature, clients and therapists view the therapeutic alliance differently (Ardito and Rabellino, 2011; Wampold and Imel, 2015). Clients place greater emphasis on collaborative working (Bachelor, 2013), and base their assessments of the relationship on their own past

experiences in similar situations (Horvath, 2000). Therapists by contrast view the relationship through a “*theoretical lens*”, assessments are based on what theory suggests is a good therapeutic relationship and what they observe during the session (Horvath, 2000, p.168; Ardito and Rabellino, 2011).

#### **6.4.2.3.2 Relationship to the process rating measures**

There were moderate correlations between total scores on the BATS and total scores on the ADAPTA PRS and AESOPS PRS with one exception; the BATS total scores correlated strongly with total frequency scores on the AESOPS PRS. The positive correlations were statistically significant. Weak correlations were found between the BATS total scores and total scores on the UKATT PRS; the positive relationships were not statistically significant. One explanation for the findings may be how the items were scored. The UKATT PRS rated item frequency, not extensiveness: “*Each time an item occurs it is marked in the space provided for frequency...*” (Middleton et al., 2001, p.4). If an item occurred once but was addressed in detail or depth, the therapist would still only receive a score of 1 ‘a little’. ADAPTA PRS and AESOPS PRS scored item frequency and extensiveness. Similar to the BATS, a higher rating was given to therapists who addressed an item in more detail or depth.

In the main, the BATS total scores did not correlate more highly with scores for selected items than total scores on the process rating measures. There was one exception; frequency scores for selected items did correlate more highly than total frequency scores on the AESOPS PRS. ADAPTA may have shown a similar pattern to AESOPS had the item ‘plan behaviour’ been included in the analyses. The findings support the notion that adherence and competence are not closely related (Perepletchikova and Kazdin, 2005) – the quality scores generally correlated less highly than the frequency scores.

Overall, the findings provide additional support for convergent validity of the BATS; stronger evidence is provided for the frequency and extensiveness ratings, rather than the quality ratings.

#### **6.4.2.4 Inter-rater reliability**

Inter-rater reliability of the item scores was examined using the weighted kappa and the ICC. Weighted kappa coefficients ranged from 0.73 to 0.96, indicating good to very good reliability between the two raters. A similar pattern was found for the ICCs; ICCs ranged

from 0.74 to 0.96, indicating good to excellent levels of reliability. The weighted kappa provided a more conservative estimate than the ICC for the majority of items, although the strength of agreement was similar for seven of the eight items. For the remaining items, the weighted kappa was equivalent to the ICC. In relation to the previous literature, the ICCs for individual items were comparable (e.g., AESOPS PRS, Watson et al., 2013a; YACS, Carroll et al., 2000), if not better than (e.g., UKATT PRS, Tober et al., 2008; IAC Treatment Fidelity Instrument, Torrey, 2012), the measures identified in Study 1.

### **6.4.3 Study strengths and limitations**

#### **6.4.3.1 Sample size**

A potential weakness of the study is that the sample size was not calculated *a priori*. A sample size calculation would have determined the number of recordings needed to avoid false-negative results (Noordzij et al., 2010)<sup>85</sup>. Tests with smaller samples may give results that are not sufficiently powered to detect “*important effects which are truly present in a population*” (Bowers et al., 2014, p.61). In such cases, researchers may fail to reject the null hypothesis, leading to a type II error. Sample sizes for this study were based on: i) previous work exploring psychometric properties within the addiction field (e.g., Lane et al., 2005; Torrey, 2012; Watson et al., 2015), and ii) what was feasible to do within the constraints of the project. CIs for the reliability and validity estimates were computed, providing an alternative means of evaluating the precision of the findings. Based on the findings, it was concluded that the BATS demonstrated acceptable levels of IRR among the two raters. Appendix E shows how the sample size could have been calculated in a hypothetical study. While *post hoc* calculations are not recommended (Walters, 2009), the required sample size for testing IRR in the hypothetical study ( $n=21$ ) was comparable to the number of recordings selected for double rating in the current study ( $n=20$ ).

That emphasis was given to the number of sessions required for testing inter-rater reliability could also be seen as a weakness. In Study 1, IRR was the most commonly reported test for examining the psychometric properties of the identified measures. As such, priority was given to the IRR analyses in the current study. This emphasis may have impacted on the convergent validity analyses. Indeed, the CIs for the correlation estimates were all relatively wide, indicating some uncertainty about the magnitude of the true

---

<sup>85</sup> The required sample size for a given power is dependent on the chosen significance level ( $\alpha$ ) and the smallest clinically important effect (Bowers, 2014).

coefficient (Bowers et al., 2014). This is perhaps unsurprising given the chosen correlation coefficient; non-parametric tests require larger samples to achieve similar levels of power than their parametric counterparts (Bland, 2017). A larger sample size may have increased the precision of the estimates for the population (Bowling, 2014). That being said, the results do provide consistent support for convergent validity of the BATS in terms of the WAI, and two of the fidelity measures (ADAPTA PRS, and AESOPS PRS). The most compelling evidence was for the relationships between the BATS and the WAI. In the three RCTS, the WAI short form was rated from the perspective of the client and the therapist. Unlike the BATS, the WAI was not an observer rating scale, which makes the correlations all the more impressive.

#### **6.4.3.2 Weighted kappa versus the ICC**

A key strength of this study is that the weighted kappa was used to examine IRR. Weighted kappa provides a good indication of agreement when using ordinal data with three or more categories of response (Watson and Petrie, 2010). The estimate gives credit to near misses by adjusting for the degree of disagreement between categories (Bowers et al., 2014). In the current study, the weighted kappa was considered preferable to the chosen form of ICC for three reasons. First, the ICC is not appropriate for ordinal measurement; the estimates were generally higher than the weight kappa. Second, mixed-effects models cannot be generalised (unlike the weighted kappa) to other raters; the results represent the reliability of the specific raters involved in the study (Koo and Li, 2016; Trevethan, 2017). Lastly, the ICC for a single rater is sensitive to the number of raters; a large number of raters will produce narrower confidence intervals for the ICC estimate than a small number of raters (Landers, 2015).

However, Cohen's kappa (similar to the weighted kappa but for two categories) has two notable limitations. First, it is "*sensitive to the proportion of subjects in each category*" (Bowers, 2014, p.264). For example, if two studies have the same proportion of observed agreement across two categories, the maximum value of kappa will occur in the study where the prevalence, i.e., the relative number in each category, is closer to 50% (Mandrekar, 2011). Second, it is susceptible to bias between the raters – the frequency at which raters choose a particular category differs (Streiner et al., 2015). As the differences between raters (bias) increases so does the value of kappa. These limitations are generalisable to the weight kappa (Flight and Julios, 2015). Caution is, therefore, needed when comparing the results of different studies (Mandrekar, 2011; Bowers, 2014).



The advantages of the ICC also need acknowledging (Streiner et al., 2015). For example, the ICC can deal with missing data, and account for multiple raters and multiple categories of response. Streiner et al. (2015) argue that the ICC provides “*a unifying framework that ties together different ways of measuring inter-rater agreement*” (Streiner et al., 2015, p.179). Given these advantages, it is perhaps unsurprising that the ICC was widely reported in the validation literature identified in Study 1. However, the articles reviewed varied on the information provided about the chosen ICC model. Incomplete information raises concerns about the correctness of the analyses, and makes comparisons between studies difficult (Koo and Li, 2016). There is also the issue of using parametric tests with ordinal data. The weighted kappa is interpretable as the ICC when the variability between raters contributes to the denominator of the estimate, the total variation (Fleiss and Cohen, 1973) (Appendix E). This equivalence highlights further the preference of using weight kappa for ordinal data with at least three categories.

## **6.5 Conclusion**

The BATS was developed further by investigating its psychometric properties. Specifically, the BATS was subjected to convergent and IRR analyses. Convergent validity was examined in two ways. First, relationships between the BATS and the Working Alliance Inventory (WAI) were explored. The results showed that the BATS and the WAI-S describe similar behaviours, focusing on the collaborative relationship between client and therapist, providing support for convergent validity of the BATS. Second, total scores on the BATS were compared with total scores from three process rating measures (ADAPTA PRS, AESOPS PRS, and UKATT PRS<sup>86</sup>). The relationships between the BATS and two of the measures (ADAPTA PRS and AESOPS PRS) provided additional support for convergent validity of the BATS. However, the UKATT PRS painted a different story; weak, non-significant relationships were found. One explanation for this difference is that the UKATT PRS rated item frequency, rather than item extensiveness. Similar to the BATS, ADAPTA PRS and AESOPS PRS scored item frequency and extensiveness; a higher rating was given to therapists who addressed an item in more detail or depth.

While there is evidence of convergent validity, the results need to be interpreted with

---

<sup>86</sup> ADAPTA PRS = ADAPTA (Addressing Drinking Among Patients: comparing Two Approaches) Process Rating Scale; AESOPS PRS = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale; UKATT PRS = UKATT (United Kingdom Alcohol Treatment Trial) Process Rating Scale.

some caution. The CIs for the correlation estimates were all relatively wide, indicating some uncertainty about the magnitude of the true coefficient (Bowers et al., 2014). A larger sample size may have increased the precision of the estimates for the population (Bowling, 2014); further replication is needed to ensure consistency of effect (Kazdin, 2009). That being said, there is compelling evidence for convergent validity of the BATS from both the WAI and the two fidelity measures (ADAPTA PRS and AESOPS PRS). It was concluded, therefore, that initial support for convergent validity of the BATS was provided. Inter-rater reliability of the item scores was examined using the weighted kappa and the ICC. Weighted kappa coefficients indicated good to very good reliability between the two raters. A similar pattern was found for the ICCs, which showed good to excellent levels of reliability.

The weighted kappa provided a more conservative estimate than the ICC for the majority of items. The weighted kappa was the preferred method for analysing ordinal data with more than three categories. Overall, the results showed that the BATS demonstrated acceptable levels of IRR among two raters. The next chapter pulls together each of the four studies, discussing how the BATS was developed in relation to the existing literature.

## **Chapter 7**

### **General discussion**

#### **7.1 Introduction**

The last four chapters have each described one of the four studies undertaken to develop the BATS. This chapter provides an overview of the project, summarising the findings from each of the four studies. The BATS is discussed within the context of the psychotherapy research. This includes consideration of the strengths and limitations of the overall project. This chapter concludes by highlighting the clinical implications and future directions of the BATS.

#### **7.2 Project overview**

The harms related to alcohol and other drugs are well documented. This has led to the development of effective psychological therapies. The standard for determining the effectiveness of psychological therapies is the randomised controlled trial (RCT) (Wampold, 2015). This powerful research design is advantageous because it minimises the risk of confounding (Bothwell et al., 2016). As such, treatment manuals have become common place for standardising the way therapies are delivered. The problem with treatment manuals is they do not guarantee that a therapy will be delivered as intended; to ascertain adherence to manualised therapies, treatment delivery must be assessed for fidelity (Schoenwald et al., 2011). This is important for evaluating treatment effectiveness as it enables treatment effects to be accurately attributed (Tober et al., 2008).

To be effective in routine practice, treatments shown to be effective in research trials need to be delivered with fidelity. At present, fidelity to addiction therapies is monitored during clinical supervision. This process can be problematic as it relies heavily on self-reports. It is argued that measures of fidelity will better support implementation and sustainability of evidence-based practices. Existing measures developed in the context of research trials are generally unsuitable for use in routine practice because they are either too long (e.g., Carroll et al., 2000) or focus on one specific treatment modality (e.g., Madson et al., 2005). In routine practice, therapists use a range of therapies, from brief advice to intensive

specialist treatment, to address clients' alcohol and drug use problems (Raistrick et al., 2006). Therapists will often respond flexibly drawing on techniques from different therapeutic approaches. To have utility in routine practice, a transtheoretical measure is needed to evaluate therapist delivery of evidence-based therapies for addressing drug and alcohol use problems.

The current project aimed to develop the Brief Addiction Therapist Scale (BATS), an evidence-based tool for monitoring treatment delivery in routine practice. The development of the BATS comprised four separate studies. Study 1 identified fidelity measures from the literature that evaluate therapists' delivery of psychological therapies for addressing alcohol and drug use problems. Study 2 generated items for potential inclusion in the BATS using the identified measures as a basis. Consideration was also given to how the items should be scored. Study 3 obtained a consensus from a group of experts on the content of the BATS, including the items and response format. Study 4 developed the BATS further by testing its psychometric properties.

### **7.3 Summary of the findings**

This project developed a transtheoretical measure to monitor the delivery of widely used therapies in addiction (i.e., the BATS). Initial psychometric analysis of the scale indicated that it is psychometrically sound in terms of: inter-rater reliability, face validity, and convergent validity. A more detailed summary of the findings from each of the four studies is provided below.

#### **7.3.1 Study 1: Identifying existing measures**

A literature review was conducted to identify fidelity measures that evaluate the delivery of psychological therapies used for alcohol and drug use problems. Twenty six measures were identified, covering 26 different therapies. Most of the measures were developed for use in the addiction field; although six measures focused on mental health (e.g., depression), and five related to a particular therapeutic approach (e.g., cognitive therapy). The majority of the measures were developed solely for research purposes. In addition to research, measures were designed for use in other contexts, such as, training, and supervision. Typically the measures evaluated treatment adherence and/or therapist competence in delivering one particular modality, with cognitive-behavioural therapy being the most common. The majority of measures focused on therapies widely used in addiction (e.g., motivational interviewing). The remainder of measures were aligned with

therapies not commonly used in addiction; however, these measures had been adapted for treating alcohol and/or drug use problems (e.g., supportive-expressive therapy).

The number of items on each of the identified measures ranged from 11 to 96, with a mean of 33 items – the measures developed solely for use in research were generally longer than those developed for other contexts. Items were scored using four main response formats: adjectival scales, Likert scales, ordered-categorical scales, and dichotomous scoring<sup>87</sup>. Adjectival scales were most frequently used to rate treatment adherence, and Likert scales to rate therapist competence. Generally, items were scored using 5-point and 7-point scales. A ‘not applicable’ option was provided in some measures, typically for rating therapist competence when the specified behaviour was not observed during the session. A range of analytic methods were used to support reliability and validity of the measures. The measures identified from the review formed the basis of generating items for the BATS<sup>88</sup>.

### **7.3.2 Study 2: Generating an item pool**

#### **7.3.2.1 Items for the BATS**

Items from the identified measures were analysed using a form of thematic analysis (Braun and Clarke, 2006); items were grouped based on what aspect of therapeutic practice they targeted. Thirty three themes were developed and grouped into five meta-themes. The first meta-theme, session management, focused on the techniques therapists may use to manage the therapy session. The second meta-theme, medication and case management, focused on therapists’ discussion of the clients’ medication, and involvement in self-help groups and other services. The third meta-theme, interventions to increase awareness, concentrated on the techniques therapists may use to increase clients’ understanding of their behaviours, thoughts, feelings and relationships. The fourth meta-theme, interventions to change behaviour, related to the techniques therapists may use to help clients change their behaviour and achieve their treatment goals. The last meta-theme, core skills, related to therapists’ style, how therapists delivered the session.

---

<sup>87</sup> Adjectival scales are ordinal scales with unipolar response options (e.g., ‘not at all’ to ‘extensively’). Likert scales are similar to adjectival scales but comprise bipolar response options (e.g., ‘very poor’ to ‘very well’). Ordered-categorical scales are Likert-type scales that contain item-specific descriptive anchors. Dichotomous scoring includes binary response options (usually ‘yes’ and ‘no’).

<sup>88</sup> In addition to identifying relevant fidelity measures, the literature review also considered the methods used by the original authors to examine reliability and validity of the measures. This was important for informing the psychometric work on the BATS.

The thematic structure was refined in collaboration with the supervision team, in order to identify the themes most relevant to the BATS. Themes concerning therapy specific techniques were removed as the BATS was designed to be transtheoretical; for example, 'session content' (grouped in meta-theme 1 'session management') was removed, as the coded items focused on the particular therapeutic approach or the specific techniques delivered during the session. The number of themes was reduced from 33 to 18. Exemplar items were chosen from each of the remaining 18 themes. Most of the exemplars were adapted from the extracted items on the identified measures; eight were newly constructed but reflected aspects of practice targeted in the identified measures. The 18 exemplar items were chosen to reflect the key features of therapies widely used in the treatment of alcohol and drug use. The 18 items were reviewed in Study 3 'agreeing the content' for potential inclusion in the BATS.

### **7.3.2.2 Response formats for the BATS**

At this stage, consideration was given to scoring – how items on the BATS should be scored. Response formats were generated using the identified measures as a basis. Two formats were generated: the first measured extensiveness, and the second measured quality. Most of the measures identified in Study 1 asked raters to score: i) the extent to which therapists carried out item specific behaviours (treatment adherence), and ii) the quality with which therapists performed the behaviours (therapist competence).

### **7.3.3 Study 3: Agreeing the content**

A modified three-round Delphi survey was used to obtain a consensus among a group of experts in the fields of addiction and psychotherapy on the content of the BATS (i.e., the items and response format). Of the 19 experts invited to take part, 12 participants took part in round one, 12 in round two, and 10 in round three. Participants were from Europe and North America, and each had over 15 years' experience in the areas of addiction and/or psychotherapy. Most participants had a dual work role, with both clinical and academic components. The diversity of participants' backgrounds assured a wide base of knowledge and expertise (Powell, 2003).

#### **7.3.3.1 Items for the BATS**

In round one, participants rated the 18 items generated in Study 2 on importance and comprehensibility. Of the 18 items, 11 achieved group consensus for both dimensions. Three of the items achieved consensus for importance only; participants agreed that the

items were important for inclusion in the BATS but considered the wording unclear. The remaining four items were removed for not meeting the study criteria (i.e., a median rating between 5 and 7<sup>89</sup>, an IQR of 2 or less, and consistency between participants' qualitative feedback and item ratings). Because comprehensibility was an issue across items, all 14 items were revised to improve clarity.

In round two, participants rated the revised 14 items on importance and comprehensibility. The item pool was reduced by another two items, as they did not meet the study criteria. Based on participants' feedback, five items were retained (with no changes) and seven were revised to improve comprehensibility. At this stage, consideration was given to how items on the BATS should be scored. Two response formats were proposed: extensiveness, and quality. Participants rated the extent to which they agreed these methods of rating were appropriate for use in the BATS. Ratings were similar for both response formats. The extensiveness scale was chosen because participants' feedback suggested the extensiveness scale would be easier and more reliable to rate than the quality scale.

In the final round, participants rated 12 items on importance only. Item comprehensibility was not rated, information provided in previous rounds was considered sufficient for amending item wording. All 12 items achieved consensus: the results showed a high level of group agreement, and provided evidence of convergence of participants' opinions. Participants' feedback highlighted the importance of including item definitions and scoring instructions, particularly for items not considered relevant for every session. Suggestions for rewording items were given. In total, nine items were retained unchanged, and three items were revised. These 12 items were included in the BATS.

### ***7.3.3.2 Response formats for the BATS***

In round two, participants were asked to consider how items included in the BATS should be rated. Two scoring methods were proposed: extensiveness and quality. Participants' ratings were similar for both methods. The extensiveness scale was chosen because participants' feedback suggested the extensiveness scale would be easier and more reliable to rate than the quality scale. On completing Study 3, the first version of the BATS was developed; this included the 12 items and the extensiveness response format.

---

<sup>89</sup> On the Likert scale: 5 = agree somewhat, and 7 = strongly agree.

#### **7.3.4 Study 4: Testing reliability and validity**

The BATS was developed further by investigating its psychometric properties. Specifically, the BATS was subjected to convergent validity and inter-rater reliability analyses.

Convergent validity was examined in two ways. First, relationships between the BATS and the Working Alliance Inventory (WAI) were explored. There were moderate, positive correlations between total scores on the BATS and total scores on the WAI short form (WAI-S) for both client and therapist. These correlations were statistically significant. The results show that the BATS and the WAI-S describe similar behaviours, focusing on the collaborative relationship between client and therapist. Thus, providing support for convergent validity of the BATS. Second, total scores on the BATS were compared with the total scores of three fidelity measures (ADAPTA PRS, AESOPS PRS, and UKATT PRS<sup>90</sup>). There were moderate to strong correlations between the BATS and the ADAPTA PRS and AESOPS PRS. These correlations were statistically significant, providing further evidence of convergent validity. However, the UKATT PRS painted a different story. Weak, non-significant correlations were found between the BATS and the UKATT PRS. One explanation for this different picture is that the UKATT PRS rated item frequency, rather than item extensiveness. If an item behaviour occurred once but was addressed in detail or depth, the therapist would still only receive a score of 1 'a little'. ADAPTA PRS and AESOPS PRS, by contrast, scored item frequency and extensiveness; similar to the BATS, a higher rating was given to therapists who addressed an item in more detail or depth.

While there is evidence of convergent validity, the results should be interpreted with some caution. The CIs for the correlation estimates were all relatively wide, indicating some uncertainty about the magnitude of the true coefficient (Bowers et al., 2014). A larger sample size may have increased the precision of the estimates for the population (Bowling, 2014); further replication is needed to ensure consistency of effect (Kazdin, 2009). That being said, there is compelling evidence for convergent validity of the BATS from both the WAI and the two fidelity measures (ADAPTA PRS and AESOPS PRS). It was concluded, therefore, that initial support for convergent validity of the BATS was provided.

---

<sup>90</sup> ADAPTA PRS = ADAPTA (Addressing Drinking Among Patients: comparing Two Approaches) Process Rating Scale; AESOPS PRS = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale; UKATT PRS = UKATT (United Kingdom Alcohol Treatment Trial) Process Rating Scale.



Inter-rater reliability of the item scores was examined using the weighted kappa and the ICC. Weighted kappa coefficients ranged from 0.73 to 0.96, indicating good to very good reliability between the two raters. A similar pattern was found for the ICCs; ICCs ranged 0.74 to 0.96, indicating good to excellent levels of reliability. The weighted kappa provided a more conservative estimate than the ICC for the majority of items, although the strength of agreement was similar for seven of the eight items. For the remaining items, the weighted kappa was equivalent to the ICC. Overall, the results showed that the BATS demonstrated acceptable levels of IRR among two raters.

#### **7.4 The BATS within the context of psychotherapy research**

There is considerable evidence for the effectiveness of psychological therapies among adults with alcohol and drug use problems (Moos, 2007). The rigour of RCTs and meta-analyses has increased, thereby achieving “*the status of gold standard for therapeutic evidence*” (Bothwell et al., 2016, p.2179). One reason for the ‘gold standard’ status is that they are sufficiently powered to demonstrate the effectiveness of treatments (Wampold and Imel, 2015). This has replaced earlier scepticism about the ability of psychological therapies to improve the outcomes of clients with alcohol and drug use problems (Longabaugh et al., 2005).

Two of the most notable RCTs in the addiction field are Project MATCH (Matching Alcohol Treatments to Client Homogeneity; Project MATCH Research Group, 1997) and UKATT (United Kingdom Alcohol Treatment Trial; UKATT Research Team, 2005). Project MATCH is the largest psychotherapy outcome study. The multicentre trial, conducted in the United States, involved 1726 participants and compared twelve-step facilitation, cognitive-behavioural therapy, and motivational enhancement therapy (MET). UKATT was “*a British version of Project MATCH*” (Green and Latchford, 2012, p.17), and investigated two approaches for treating alcohol dependence. The multicentre trial involved 742 participants, and compared MET with a more intensive treatment called social behaviour and network therapy. What is interesting about both RCTs, and others like them, is that the results showed no significant differences between treatments – all were equally effective in terms of treatment outcome. The broad equivalence of psychotherapies reflects both the addiction (e.g., Imel et al., 2008) and psychotherapy (e.g., Wampold et al., 1997) literature.

Despite the many effectiveness studies, we know very little about why therapy works. It has been hypothesised that specific ingredients underlie the effectiveness of psychological therapies (Wampold and Imel, 2015). This means that therapy purportedly works through the interventions and processes delivered by therapists that are specific to the theoretical orientation of the treatment. For example, cognitive-behavioural therapy (CBT) is based on models from cognitive and behavioural psychology (Moos, 2007). As such, CBT is thought to work by improving the clients' ability to cope with high-risk situations (any situation that is likely to precipitate substance use), thereby enhancing the clients' self-efficacy to abstain from substance using behaviours (Magill et al., 2015; Carroll, 1997). While plausible in theory, current evidence provides weak support for the specific effects of treatments (Wampold and Imel, 2015). Following on from the CBT example, a review by Morgenstern and Longabaugh (2000) found little evidence for the hypothesised ingredients of CBT, concluding that *"research has not yet established why CBT is an effective treatment for alcohol dependence"* (p.1475). The failure of studies to establish the specific effects of treatment is not *"unique"* to CBT or addiction (Longabaugh, 2010, p.2128). The findings are consistent with the wider literature (Wampold and Imel, 2015) – irrespective of research design (e.g., Bell et al., 2013), psychotherapy (e.g., Apodaca and Longabaugh, 2009), or target problem (e.g., Kazdin and Nock, 2003).

Investigating the specific effects of psychotherapy is difficult (Wampold and Imel, 2015); *"negative findings may reflect methodological flaws of prior studies"* (Morgenstern and Longabaugh, 2000, p.1475). More sophisticated methods have allowed for progress to be made in this area (Magill et al., 2015). For example, Kiluk et al. (2010) showed that observer-rated quality of coping responses mediated in part the effects (duration of abstinence from treatment termination) of CBT over treatment as usual for substance use problems. Unlike previous studies, the authors used: i) a *"superior index"* for explaining the relationship between CBT and client outcomes – prior studies focused solely on quantity of behaviour, ii) used more robust analytic methods, and iii) standardised treatment delivery through a computerised form of CBT (Longabaugh, 2010, p.2128). While this example is notable, the findings need to be interpreted with caution, as the sample size was relatively small ( $n=52$ ); further replication is needed to ensure consistency of effect (Kazdin, 2009).

At present, there is more compelling evidence for the common factors of psychotherapy – *"therapeutic elements that are common to all or most psychotherapies"* (Wampold, 2015, p.270). Several meta-analyses point to the operation of common factors in determining

positive treatment outcomes (Wampold and Imel, 2015). For example, Elliott et al. (2011) conducted a meta-analysis examining the importance of the variable 'empathy'. The meta-analysis, involving 57 studies and 3,599 clients, found a moderate correlation between ratings of therapist empathy and treatment outcome. It is, therefore, unsurprising that factors related to empathy have been identified as important in psychotherapy (Wampold, 2015). These factors include: positive regard and affirmation (Farber and Doolin, 2011), and congruence and genuineness (Kolden et al., 2011). The majority of meta-analyses have not examined the potential impact of confounding. For example, client characteristics may explain the identified relationship between empathy and therapy outcome:

*"It is clearly easier for a therapist to be warm and caring toward a motivated, disclosing and cooperative patient than to one who is interpersonally aggressive, and the former types of patients will most likely have better outcomes than the latter" (Wampold, 2015, p.273)*

The therapeutic alliance is the most researched common factor. It is the only factor where studies have started to address the potential confounders of the alliance-outcome relationship (Wampold and Imel, 2015). For example, Horvath et al. (2011) conducted a comprehensive meta-analysis, aggregating data from 190 studies, and found a moderate association between the alliance and treatment outcome. The study was comprehensive because it also examined the impact of several variables that potentially moderated the effect of the alliance. Some of the moderators included: the alliance measure used, the alliance rater (perspective of the client, therapist, or observer), and the treatment modality. The results showed that the potential confounders did not threaten the conclusion of the alliance (and the therapists' contribution) as an important therapeutic factor. Similar to empathy, there is evidence to suggest that constructs related to the alliance are important for the provision of effective treatment delivery. These factors, labelled goal consensus and collaboration, highlight the importance of therapists working together with clients towards achieving the clients' treatment goals (Tryon and Winograd, 2011).

The research summarised so far has attempted to explain why effective therapies work. Two possible explanations have been presented. The first explanation is that specific ingredients of treatments are responsible for therapeutic change. These effects are underpinned by the specific theoretical approaches of the treatments. Therapies may work in *"very different ways, but each of these ways is equally effective"* (Green and Latchford,

2012, p.19). The second explanation is that there are factors common in all psychotherapies. These factors are not specific to a particular treatment model. Therapies may work because they “*all share the same active ingredients*” (Green and Latchford, 2012, p.19). Given the evidence to date, it is not unreasonable to think that both theories have value. That is, there are specific contributions of any one particular treatment model, as well as a common set of factors that cuts across all effective therapies. The BATS reflects this dual idea.

Items on the BATS are clearly consistent with the identified common factors of psychotherapy. For example, item 2 ‘collaboration’ focuses on the extent to which therapists attempted to work together with the client. However, the BATS also includes therapist techniques which have been popularised in the context of a particular theoretical approach, but have been incorporated into the practice of many treatment models (Petrik et al., 2013). For example, item 11 ‘behaviour change planning’ considers the extent to which therapists enabled a plan for changing the clients’ substance use, or maintaining change, to be discussed. The development of the BATS pools together what we do know about the active ingredients of therapy, both from the literature and from expert opinion. The BATS has demonstrated reliability and validity for evaluating treatment delivery in the addiction field.

## **7.5 Strengths and limitations**

Strengths and limitations of each study have been described in the corresponding chapters. This section will consider overarching strengths and limitations of the overall project. The multimethod design used to develop the BATS is a key strength of the current project. Using different methods is advantageous because each one offers a different perspective on the content of the scale. This approach provided a means of testing and improving our imperfect view of reality (Yardley and Marks, 2004). The four interrelated studies were designed and conducted to address specific research questions (Morse, 2003). Driven by the overarching aim (to develop the BATS), the studies provided complementary information that compensated for the shortcomings of using only one method (Davis et al., 2011; Morse, 2003). For example, the item pool generated in Study 2 was shaped, to an extent, according to the predispositions and biases of the researcher and the supervision team (Patton, 2002). The inclusion of the Delphi exercise in Study 3 offset this limitation by providing a complementary perspective; selected experts in the fields of addiction and psychotherapy agreed on the content of the BATS (i.e., the items

and response format). The combination of methods improved on other scale development studies in the field (e.g., Barber et al., 1996; Davidson et al., 2004), and ensured that the BATS was developed as comprehensively as possible (Morse, 2003).

A further strength is that all four studies were conducted using rigorous and well defined methods. The use of rigorous methods throughout the project supports the development of a more robust measure. For example, the literature was comprehensively searched in Study 1 to identify fidelity measures that evaluate the delivery of psychological therapies for alcohol and drug use problems. The review used a systematic approach, and efforts were made to minimise the risk of selection and publication biases. This robust approach enhanced the degree of confidence in the results. When taken as a whole, the four studies offer compelling evidence that the BATS is a reliable and valid tool for evaluating the range of widely used therapies in addiction. Given the emphasis on method, the project allows for the future replication and verification of results. In other words, the project produced a 'positive' outcome that may be objectively viewed by peers.

The supervisors' contribution to the development of the BATS was a strength and a limitation of the project. In each of the four studies, efforts were made to enhance the quality of analysis. For example, in study 2, the item pool was generated in collaboration with the supervision team, reducing the potential bias that comes from a single researcher (Patton, 2002); alternative explanations and organising schemes were considered, and the credibility of the findings was enhanced (Patton, 2002; Nowell et al., 2017). However, this means that the development of the BATS was shaped according to the experience and professional backgrounds of the researcher (HC) and the supervision team (GL, BB, and GT). HC has a background in psychology and applied health research, having worked as a researcher across addiction and mental health fields. GL, a registered clinical and health psychologist, is the joint director for the Doctor of Clinical Psychology programme at Leeds. He has a range of research interests, including psychotherapy (with a particular interest in motivational interviewing), and the process of change. BB is a Chartered Psychologist and an Associate Professor in Psychological Health and Wellbeing. Her portfolio of research focuses around monitoring, managing, and modifying mental health and well-being. GT was, until recently, employed by a NHS mental health trust as a consultant addiction psychologist, practising, teaching and supervising in the delivery of addiction treatment. Her research is in the measurement of addiction treatment delivery, and the effectiveness of addiction treatment and the nature and measurement of substance dependence. Given

the backgrounds of the researcher and the supervision team, there was a propensity to foreground factors relating to cognitive and motivational treatments, rather than those from other treatment modalities, such as psychodynamic therapy. While it is impossible to be value free or to attain objectivity, being open and reflexive about our personal and intellectual biases helps to position the research and enhance the credibility of the findings (Mays and Pope, 2000; Willig, 2013).

One potential weakness of the project is basing the BATS on the existing literature. At present, there is an emphasis in the literature on cognitive-behavioural and motivational therapies, with few exemplars of psychodynamic treatments (Wampold and Imel, 2015). One explanation for this is that psychodynamic treatments are more difficult to manualise than other forms of therapy (Wampold and Imel, 2015). As Shedler (2010) summarised, the therapy processes in many psychodynamic trials have been "*inadequately specified and monitored*" (p.106). Given the current emphasis in the literature, the BATS potentially excludes key features of psychodynamic therapies. At present, the potential exclusion of key items is not considered problematic. In routine practice, therapists are encouraged to deliver evidence-based treatments for addressing alcohol and drug use problems (Raistrick et al., 2006). Clinical guidelines, in particular, advocate the use of manualised therapies because they have the best available supportive evidence (Green and Latchford, 2012). For example, guidelines for alcohol-use disorders (harmful drinking and alcohol dependence) state that: "*Psychological interventions should be based on a relevant evidence-based treatment manual, which should guide the structure and duration of the intervention*" (National Institute for Health and Clinical Excellence, 2011, p.10). Given this emphasis on manualised therapies, psychodynamic treatments are not routinely used in the addiction field. To have real-world application, the BATS needed to reflect common practice. It was important, therefore, to base the BATS on the existing literature. By doing so, the BATS has the potential to support the implementation and sustainability of evidence-based practices.

Lastly, that items on the BATS are not always relevant for every session could also be seen as a weakness. For example, item 8 'homework reviewed' is not generally applicable for clients' attending their first session. The concern is that the BATS may not be appropriate for targeting different stages of the therapy process, as the content of an initial assessment differs to later sessions where the focus is on planning or maintaining change. Herein lay the challenge, to develop a 'one size fits all' scale relevant for different therapies and

different stages of the therapy process. To address this concern, focus was given to the overall presentation of the BATS. Items were grouped based on the different stages of change: i) items considered relevant to most sessions irrespective of clients' readiness to change, ii) items applicable for building motivation for change, and iii) items appropriate for planning or maintaining change. This solution, while not ideal, was considered viable to ensure the BATS was suitable for use in routine practice. An alternative option would have been to design multiple scales, each targeting a different stage of therapy. Such an option would have increased the length and complexity of the BATS reducing its utility in routine practice. Support for the 'one size fits all' solution was provided in Study 4; the selected recordings of therapy sessions included clients at different stages of therapy.

## **7.6 Clinical implications and future directions**

The BATS has real world application. The project developed a brief, evidence-based tool for monitoring and evaluating the delivery of psychological therapies used in routine practice. The BATS provides a useful tool for training and supervision, which has the potential to impact on therapist competence and treatment outcomes. There is compelling evidence for the utility of the BATS in routine practice. For example, a NHS addiction service in England has already incorporated the scale to support peer supervision. Permission to use the BATS has also been given to addiction services in Estonia and Wales. Future studies would benefit from exploring further the implementation of the BATS in clinical services, focusing on the provision of effective treatment delivery. Preliminary work is being undertaken to examine the relationship between the BATS and treatment outcome, highlighting the scale's value in facilitating future process research. Such research may be particularly useful in developing our knowledge of the active ingredients of effective treatments for alcohol and drug use problems.

This project provides initial support for reliability and validity of the BATS. The scale would benefit from future studies testing further its psychometric properties. There are two ways future studies could approach this work. The first approach is to build on the methods used in Study 4. For example, inter-rater reliability analyses were based on the scores of two raters who had extensive experience of fidelity measurement. Exploring the reliability of the BATS with a larger group of therapists, who are not experts in fidelity measurement, would broaden the conclusions drawn. The second approach is to test other forms of reliability and validity that were not examined in Study 4 but would be useful in the ongoing development of the BATS. For example, test retest reliability to assess the

consistency of the BATS scores over time, and construct validity to explore the underlying constructs of the BATS. Irrespective of the approach, future studies would benefit from using larger samples (derived from *a priori* sample size calculations) to increase the precision of the results (Bowling, 2014). To this end, it would be useful to focus on one particular therapy type (e.g., therapy sessions delivered in routine practice settings), rather than the range of therapies used in Study 4<sup>91</sup>. If validated in other studies, researchers and service providers might use adaptations of the BATS to save time and resource on developing another similar measure. Such studies would, therefore, support the applicability of the BATS to other treatments and populations.

## **7.7 Conclusion**

This thesis has described the development of the Brief Addiction Therapist Scale (BATS): an evidence-based tool for monitoring and evaluating therapists' delivery of psychological therapies used in routine practice for alcohol and drug use problems. The BATS is transtheoretical, applicable to the range of widely used therapies in addiction. It is being used to support peer supervision at a NHS addiction service, demonstrating its utility in routine practice. The development of the BATS contributes to our understanding of why therapy works, both from the literature and from expert opinion. The BATS has demonstrated acceptable levels of inter-rater reliability, and initial support for convergent validity. Preliminary work is being undertaken to examine the relationship between the BATS and treatment outcome. The BATS would benefit from future studies testing further its psychometric properties. However, the BATS clearly has real world application. The BATS provides a useful tool for training and supervision, which has the potential to impact on therapist competence and treatment outcomes.

---

<sup>91</sup> Seven therapy types were included in Study 4 'testing reliability and validity': an alcohol focused intervention (AF), a healthy living intervention (HF), behaviour change counselling (BCC), brief advice (BA), motivational enhancement therapy (MET), social behaviour and network therapy (SBNT), and routine practice therapy sessions (RP).



## References

- Aarons, G.A., Hurlburt, M. and Horwitz, S.M. 2011. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*. **38**(1), pp.4-23.
- Alkharusi, H. 2012. Generalizability theory: An analysis of variance approach to measurement problems in educational assessment. *Journal of Studies in Education*. **2**(1), pp.184-196.
- Allan, R. and Ungar, M. 2014. Developing a measure of fidelity for an ecological approach to family therapy. *Journal of Family Psychotherapy*. **25**(1), pp.26-41.
- Altman, D.G. 1999. *Practical Statistics for Medical Research*. London: Chapman and Hall/CRC.
- Altman, D.G., Machin, D., Bryant, T.N. and Gardner, M.J. 2000. *Statistics with Confidence (Second Edition)*. London: BMJ.
- Alvarez-Jimenez, M., Wade, D., Cotton, S., Gee, D., Pearce, T., Crisp, K., McGorry, P.D. and Gleeson, J.F. 2008. Enhancing treatment fidelity in psychotherapy research: Novel approach to measure the components of cognitive behavioural therapy for relapse prevention in first-episode psychosis. *Australian and New Zealand Journal of Psychiatry*. **42**(12), pp.1013-1020.
- Amini, R.L. and Woolley, S.R. 2011. First-session competency: The brief strategic therapy scale-1. *Journal of Marital and Family Therapy*. **37**(2), pp.209-222.
- Apodaca, T.R. and Longabaugh, R. 2009. Mechanisms of change in motivational interviewing: A review and preliminary evaluation of the evidence. *Addiction*. **104**(5), pp.705-715.
- Ardito, R.B. and Rabellino, D. 2011. Therapeutic alliance and outcome of psychotherapy: Historical excursus, measurements, and prospects for research. *Frontiers in Psychology*. **2**, p270.
- Arksey, H. and O'Malley, L. 2005. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*. **8**(1), pp.19-32.
- Arndta, S., Turvey, C. and Andreasen, N.C. 1999. Correlating and predicting psychiatric symptom ratings: Spearman's r versus Kendall's tau correlation. *Journal of Psychiatric Research*. **33**(2), pp.97-104.
- Bachelor, A. 2013. Clients' and therapists' views of the therapeutic alliance: Similarities, differences and relationship to therapy outcome. *Clinical Psychology and Psychotherapy*. **20**(2), pp.118-135.
- Barber, J.P. 1988. *The Penn Adherence/Competence Scale for Supportive-Expressive Psychotherapy*. Unpublished.
- Barber, J.P. 1997. *Adherence/Competence Scale for Supportive-Expressive therapy for Cocaine Dependence (ACS-SEC)*. Unpublished.
- Barber, J.P. and Crits-Christoph, P. 1996. Development of a therapist adherence/competence rating scale for supportive-expressive dynamic psychotherapy: A preliminary report. *Psychotherapy Research*. **6**(2), pp.81-94.
- Barber, J.P., Krakauer, I., Calvo, N., Badgio, P.C. and Faude, J. 1997. Measuring adherence and competence of dynamic therapists in the treatment of cocaine dependence. *Journal of Psychotherapy Practice and Research*. **6**(1), pp.12-24.
- Barber, J.P., Liese, B.S. and Abrams, M.J. 2003. Development of the cognitive therapy adherence and competence scale. *Psychotherapy Research*. **13**(2), pp.205-221.

- Barber, J.P., Mercer, D., Krakauer, I. and Calvo, N. 1996. Development of an adherence/competence rating scale for individual drug counseling. *Drug and Alcohol Dependence*. **43**(3), pp.125-132.
- Barkham, M., Firth-cozens, J. and Shapiro, D.A. 1989. Change in prescriptive versus exploratory therapy: Older clients' responses to therapy. *Counselling Psychology Quarterly*. **2**(4), pp.395-403.
- Barrowclough, C., Haddock, G., Wykes, T., Beardmore, R., Conrod, P., Craig, T., Davis, L., Dunn, G., Eisner, E., Lewis, S., Moring, J., Steel, C. and Tarrrier, N. 2010. Integrated motivational interviewing and cognitive behavioural therapy for people with psychosis and comorbid substance misuse: Randomised controlled trial. *British Medical Journal*. [Online]. **341**, pp.1-12. [Accessed 12/09/2018]. Available from: <https://doi.org/10.1136/bmj.c6325>
- Barton, S., Armstrong, P., Freeston, M. and Twaddle, V. 2008. Early intervention for adults at high risk of recurrent/chronic depression: Cognitive model and clinical case series. *Behavioural and Cognitive Psychotherapy*. **36**(3), pp.263-282.
- Bassett, S.S., Stein, L.A.R., Rossi, J.S. and Martin, R.A. 2016. Evaluating measures of fidelity for substance abuse group treatment with incarcerated adolescents. *Journal of Substance Abuse Treatment*. **66**, pp.9-15.
- Beck, A.T., Rush, A.J., Shaw, B.F. and Emery, G. 1979. *Cognitive Therapy of Depression*. New York: Guilford Press.
- Bell, E.C., Marcus, D.K. and Goodlad, J.K. 2013. Are the parts as good as the whole? A meta-analysis of component treatment studies. *Journal of Consulting and Clinical Psychology*. **81**(4), pp.722-736.
- Bigby, M. and Williams, H. 2003. Appraising systematic reviews and meta-analyses. *Archives of Dermatology*. **139**(6), pp.795-798.
- Bishara, A. and Hittner, J. 2012. Testing the significance of a correlation with nonnormal data: Comparison of pearson, spearman, transformation, and resampling approaches. *Psychological Methods*. **17**(3), pp.399-417.
- Black, N., Murphy, M., Lamping, D., McKee, M., Sanderson, C., Askham, J. and Marteau, T. 1999. Consensus development methods: a review of best practice in creating clinical guidelines. *Journal of Health Services Research and Policy*. **4**(4), pp.236-248.
- Blackburn, I.-M., James, I.A., Milne, D.L., Baker, C., Standart, S., Garland, A. and Reichelt, F.K. 2001a. The revised cognitive therapy scale (CTS-R): Psychometric properties. *Behavioural and cognitive psychotherapy*. **29**(4), pp.431-446.
- Blackburn, I.M., James, I.A., Milne, D.L. and Reichelt, F.K. 2001b. *Cognitive Therapy Scale - Revised (CTS-R)*. [Online]. [Accessed 12/09/2018]. Available from: <http://ebbp.org/resources/CTS-R.pdf>
- Blagys, M.D. and Hilsenroth, M.J. 2000. Distinctive features of short-term psychodynamic-interpersonal psychotherapy: A review of the comparative psychotherapy process literature. *Clinical Psychology: Science and Practice*. **7**(2), pp.167-188.
- Blagys, M.D. and Hilsenroth, M.J. 2002. Distinctive activities of cognitive-behavioral therapy: A review of the comparative psychotherapy process literature. *Clinical Psychology Review*. **22**(5), pp.671-706.
- Bland, M. 2000. *An Introduction to Medical Statistics (Third Edition)*. Oxford: Oxford University Press.
- Bland, M. 2017. *An Introduction to Medical Statistics (Fourth Edition)*. Oxford: Oxford University Press.
- Boswell, J.F., Gallagher, M.W., Sauer-Zavala, S.E., Bullis, J., Gorman, J.M., Shear, M.K., Woods, S. and Barlow, D.H. 2013. Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology*. **81**(3), pp.443-454.

- Bothwell, L.E., Greene, J.A., Podolsky, S.H. and Jones, D.S. 2016. Assessing the gold standard: Lessons from the history of RCTs. *New England Journal of Medicine*. **374**(22), pp.2175-2181.
- Bowers, D. 2014. *Medical Statistics from Scratch: An Introduction for Health Professionals (Third Edition)*. Chichester: Wiley.
- Bowers, D., House, A., Owens, D. and Bewick, B. 2014. *Understanding Clinical Papers (Third Edition)*. Chichester: Wiley.
- Bowling, A. 2014. *Research Methods in Health: Investigating Health and Health Services (Fourth Edition)*. Maidenhead: Open University Press.
- Brauhardt, A., de Zwaan, M., Herpertz, S., Zipfel, S., Svaldi, J., Friederich, H.C. and Hilbert, A. 2014. Therapist adherence in individual cognitive-behavioral therapy for binge-eating disorder: Assessment, course, and predictors. *Behaviour Research and Therapy*. **61**, pp.55-60.
- Brauhardt, A. and Hilbert, A. 2010. *The Adherence Control Form*. Unpublished.
- Braun, V. and Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*. **3**(2), pp.77-101.
- Breitenstein, S.M., Gross, D., Garvey, C.A., Hill, C., Fogg, L. and Resnick, B. 2010. Implementation fidelity in community-based interventions. *Research in nursing & health*. **33**(2), pp.164-173.
- Brenner, H. and Kliebsch, U. 1996. Dependence of weight kappa coefficients on the number of categories. *Epidemiology*. **7**(2), pp.199-202.
- Campbell, B.K. and Guydish, J. 2012. *Twelve Step Facilitation Adherence Competence Empathy Scale (TSF ACES) Fidelity Ratings Manual*. [Online]. [Accessed 12/09/2018]. Available from: [http://ctndisseminationslibrary.org/PDF/795\\_TSFAACES.pdf](http://ctndisseminationslibrary.org/PDF/795_TSFAACES.pdf)
- Campbell, B.K., Manuel, J.K., Manser, S.T., Peavy, K.M., Stelmokas, J., McCarty, D. and Guydish, J.R. 2013. Assessing fidelity of treatment delivery in group and individual 12-step facilitation. *Journal of Substance Abuse Treatment*. **44**(2), pp.169-176.
- Campos-Melady, M., Smith, J.E., Meyers, R.J., Godley, S.H. and Godley, M.D. 2017. The effect of therapists' adherence and competence in delivering the adolescent community reinforcement approach on client outcomes. *Psychology of Addictive Behaviors*. **31**(1), pp.117-129.
- Care Quality Commission. 2008. *Supporting Information and Guidance: Supporting Effective Clinical Supervision*. [Online]. [Accessed 12/09/2018]. Available from: [https://www.cqc.org.uk/sites/default/files/documents/20130625\\_800734\\_v1\\_00\\_supporting\\_information-effective\\_clinical\\_supervision\\_for\\_publication.pdf](https://www.cqc.org.uk/sites/default/files/documents/20130625_800734_v1_00_supporting_information-effective_clinical_supervision_for_publication.pdf)
- Carlson, K.D. and Herdman, A.O. 2012. Understanding the impact of convergent validity on research results. *Organizational Research Methods*. **15**(1), pp.17-32.
- Carroll, K.M. 1997. *Cognitive-Behavioral Coping Skills Treatment for Cocaine Dependence*. [Online]. [Accessed 12/09/2018]. Available from: [https://medicine.yale.edu/psychiatry/pdc/manuals\\_videos/CBT\\_175259\\_284\\_212\\_76\\_v1.pdf](https://medicine.yale.edu/psychiatry/pdc/manuals_videos/CBT_175259_284_212_76_v1.pdf)
- Carroll, K.M. 1998. *A Cognitive Behavioral Approach: Treating Cocaine Addiction*. National Institute on Drug Abuse Rockville, MD.
- Carroll, K.M., Connors, G.J., Cooney, N.L., DiClemente, C.C., Donovan, D.M., Kadden, R.R., Longabaugh, R.L., Rounsaville, B.J., Wirtz, P.W. and Zweben, A. 1998a. Internal validity of Project MATCH treatments: Discriminability and integrity. *Journal of Consulting & Clinical Psychology*. **66**(2), pp.290-303.
- Carroll, K.M., Martino, S. and Rounsaville, B.J. 2010. No train, no gain? *Clinical Psychology: Science and Practice*. **17**(1), pp.36-40.

- Carroll, K.M., Nich, C. and Rounsaville, B.J. 1998b. Utility of therapist session checklists to monitor delivery of coping skills treatment for cocaine abusers. *Psychotherapy Research*. **8**(3), pp.307-320.
- Carroll, K.M., Nich, C., Sifry, R.L., Nuro, K.F., Frankforter, T.L., Ball, S.A., Fenton, L. and Rounsaville, B.J. 2000. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*. **57**(3), pp.225-238.
- Carroll, K.M., O'Malley, S.S. and Nuro, K.F. 1999. *Compliance Enhancement: A Manual for the Psychopharmacotherapy of Drug Abuse and Dependence*. [Online]. [Accessed 12/09/2018]. Available from: [https://medicine.yale.edu/psychiatry/pdc/manuals\\_videos/ComplianceEnhancement\\_176455\\_284\\_21276\\_v1.pdf](https://medicine.yale.edu/psychiatry/pdc/manuals_videos/ComplianceEnhancement_176455_284_21276_v1.pdf)
- Chambless, D.C. and Hollon, S.D. 2012. Treatment validity for intervention studies. In: Cooper, H., et al. eds. *APA Handbook of Research Methods in Psychology, Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological (Volume 2)*. Washington, DC: American Psychological Association, pp.529-552.
- Chawla, N., Collin, S., Bowen, S., Hsu, S., Grow, J., Douglass, A. and Marlatt, G.A. 2010. The mindfulness-based relapse prevention adherence and competence scale: Development, interrater reliability, and validity. *Psychotherapy Research*. **20**(4), pp.388-397.
- Cicchetti, D.V. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. **6**(4), pp.284-290.
- Clark, D.M., von Consbruch, K., Hinrichs, J. and Stangier, U. 2007. *Cognitive Therapy of Social Phobia: Checklist of Therapist Competency*. Unpublished.
- Clarke, V. and Braun, V. 2013. Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The Psychologist*. [Online]. **26**(2), pp.120-123. [Accessed 12/09/2018]. Available from: <http://eprints.uwe.ac.uk/21155/3/Teaching%20thematic%20analysis%20Research%20Repository%20version.pdf>
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences (Second Edition)*. [Online]. New York: Lawrence Erlbaum Associates. [Accessed 12/09/2018]. Available from: <http://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf>
- Collins, S.E., Eck, S., Kick, E., Schröter, M., Torchalla, I. and Batra, A. 2009. Implementation of a smoking cessation treatment integrity protocol: Treatment discriminability, potency and manual adherence. *Addictive Behaviors*. **34**(5), pp.477-480.
- Connolly, M.B., Crits-Christoph, P., Shappell, S., Barber, J.P. and Luborsky, L. 1998. Therapist Interventions in Early Sessions of Brief Supportive-Expressive Psychotherapy for Depression. *The Journal of Psychotherapy Practice and Research*. **7**(4), pp.290-300.
- Cook, S., Heather, N., McCambridge, J. and United Kingdom Alcohol Treatment Trial Research Team. 2015. The role of the working alliance in treatment for alcohol problems. *Psychology of Addictive Behaviors*. **29**(2), pp.371-381.
- Copello, A., Orford, J., Hodgson, R., Tober, G. and Barrett, C. 2002. Social behaviour and network therapy: Basic principles and early experiences. *Addictive Behaviors*. **27**(3), pp.345-366.
- Costello, A.B. and Osborne, J.W. 2005. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*. [Online]. **10**(7), pp.1-9. [Accessed 12/09/2018]. Available from: <https://pareonline.net/pdf/v10n7.pdf>

- Cresswell, J.W. 2009. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (Third Edition)*. Thousand Oaks, California: Sage.
- Crits-Christoph, P., Gibbons, M.B.C., Gallop, R., Ring-Kurtz, S., Barber, J.P., Worley, M., Present, J. and Hearon, B. 2008. Supportive-expressive psychodynamic therapy for cocaine dependence: A closer look. *Psychoanalytic Psychology*. **25**(3), pp.483–498.
- Cruickshank, J. 2012. Positioning positivism, critical realism, and social constructivism in the health sciences: A philosophical orientation. *Nursing Inquiry*. **19**(1), pp.71-82.
- D'Onofrio, G., Pantaloni, M.V., Degutis, L.C., Fiellin, D.A. and O'Connor, P.G. 2005. Development and implementation of an emergency practitioner-performed brief intervention for hazardous and harmful drinkers in the emergency department. *Academic Emergency Medicine*. **12**(3), pp.249-256.
- Dalkey, N. 1969. An experimental study of group opinion: The Delphi method. *Futures*. **1**(5), pp.408-426.
- Davidson, K., Scott, J., Schmidt, U., Tata, P., Thornton, S. and Tyrer, P. 2004. Therapist competence and clinical outcome in the Prevention of Parasuicide by Manual Assisted Cognitive Behaviour Therapy Trial: the POPMACT study. *Psychological Medicine*. **34**(5), pp.855-863.
- Davis, D.F., Golicic, S.L. and Boerstler, C.N. 2011. Benefits and challenges of conducting multiple methods research in marketing. *Journal of the Academy of Marketing Science*. **39**(3), pp.467-479.
- DeCoster, J. 1998. *Overview of Factor Analysis*. [Online]. [Accessed 12/09/2018]. Available from: <http://www.stat-help.com/factor.pdf>
- Degenhardt, L., Whiteford, H.A., Ferrari, A.J., Baxter, A.J., Charlson, F.J., Hall, W.D., Freedman, G., Burstein, R., Johns, N., Engell, R.E., Flaxman, A., Murray, C.J.L. and Vos, T. 2013. Global burden of disease attributable to illicit drug use and dependence: findings from the Global Burden of Disease Study 2010. *The Lancet*. **382**(9904), pp.1564-1574.
- Delanty, G. 2005. *Social Science: Philosophical and Methodological Foundations*. Maidenhead, England: Open University Press.
- Delbecq, A.L., Van de Ven, A.H. and Gustafson, D.H. 1975. *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*. Illinois: Scott, Foresman and Company.
- Department of Health. 2017. *Drug Misuse and Dependence: UK Guidelines on Clinical Management*. [Online]. [Accessed 12/09/2018]. Available from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/673978/clinical\\_guidelines\\_2017.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673978/clinical_guidelines_2017.pdf)
- DeRubeis, R.J., Hollon, S.D., Evans, M.D. and Bemis, K.M. 1982. Can psychotherapies for depression be discriminated? A systematic investigation of cognitive therapy and interpersonal therapy. *Journal of Consulting and Clinical Psychology*. **50**(5), p744.
- DeVellis, R.F. 2017. *Scale Development: Theory and Applications (Fourth Edition)*. London: Sage.
- Diamond, I.R., Grant, R.C., Feldman, B.M., Pencharz, P.B., Ling, S.C., Moore, A.M. and Wales, P.W. 2014. Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*. **67**(4), pp.401-409.
- Donohoe, H.M. and Needham, R.D. 2008. Moving best practice forward: Delphi characteristics, advantages, potential problems, and solutions. *International Journal of Tourism Research*. **11**(5), pp.415-437.
- Donovan, D.M., Daley, D.C., Brigham, G.S., Hodgkins, C.C., Perl, H.I., Garrett, S., Doyle, S., Floyd, A.S., Knox, P.C., Botero, C., Kelly, T., Killeen, T., Hayes, C., Baumhofer, N.K.i., Seamans, C. and Zammarelli, L. 2013. Stimulant abuser groups to engage in 12-Step

- (STAGE-12): A multisite trial in the NIDA Clinical Trials Network. *Journal of Substance Abuse Treatment*. **44**(1), pp.103-114.
- du Prel, J.-B., Hommel, G., Röhrig, B. and Blettner, M. 2009. Confidence interval or p-value? Part 4 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*. **106**(19), pp.335–339.
- Elkin, I., Parloff, M.B., Hadley, S.W. and Autry, J.H. 1985. NIMH treatment of depression collaborative research program: Background and research plan. *Archives of General Psychiatry*. **42**(3), pp.305-316.
- Elliott, R., Bohart, A.C., Watson, J.C. and Greenberg, L.S. 2011. Empathy. *Psychotherapy*. **48**(1), pp.43-49.
- Erceg-Hurn, D.M. and Mirosevich, V.M. 2008. Modern robust statistical methods: An easy way to maximise the accuracy and power of your research. *American Psychologist*. **63**(7), pp.591-601.
- Farber, B.A. and Doolin, E.M. 2011. Positive regard. *Psychotherapy*. **48**(1), pp.58-64.
- Fawcett, J., Epstein, P., Fiester, S.J., Elkin, I. and Autry, J.H. 1987. Clinical management- imipramine: Placebo administration manual. NIMH treatment of depression collaborative research program. *Psychopharmacology Bulletin*. **23**(2), pp.309-324.
- Field, A. 2017. *Discovering Statistics Using IBM SPSS Statistics (Fifth Edition)*. London: SAGE.
- Fisher, D.R., Kamo, M., Sandowicz, M., Albanese, L.-A. and Beutler, L.E. 2000. *STS Therapy Rating Scale Manual- Revised*. Unpublished.
- Fleiss, J.L. and Cohen, J. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measure of reliability. *Educational and Psychological Measurement*. **33**, pp.613-619.
- Flight, L. and Julios, S.A. 2015. The disagreeable behaviour of the kappa statistic. *Pharmaceutical Statistics*. **14**(1), pp.74-78.
- Garnett, C., Crane, D., West, R., Brown, J. and Michie, S. 2015. Identification of behavior change techniques and engagement strategies to design a smartphone app to reduce alcohol consumption using a formal consensus method. *JMIR mHealth and uHealth*. [Online]. **3**(2), pp.1-9. [Accessed 12/09/2018]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4526967/>
- Gaudiano, B.A. 2008. Cognitive-behavioral therapies: Achievements and challenges. *Evidence-Based Mental Health*. **11**(1), pp.5-7.
- Ghasemi, A. and Zahediasl, S. 2012. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism*. **10**(2), pp.486-489.
- Giddens, A. 1995. *Politics, Sociology, and Social Theory: Encounters with Classical and Contemporary Social Thought*. California: Stanford University Press.
- Giraudeau, B. 1996. Negative values of the intraclass correlation coefficient are not theoretically possible. *Journal of Clinical Epidemiology*. **49**(10), pp.1205-1206.
- Godfrey, E., Chalder, T., Ridsdale, L., Seed, P. and Ogden, J. 2007. Investigating the 'active ingredients' of cognitive behaviour therapy and counselling for patients with chronic fatigue in primary care: Developing a new process measure to assess treatment fidelity and predict outcome. *British Journal of Clinical Psychology*. **46**(3), pp.253-272.
- Goldberg, D.P., Hobson, R.F., Maguire, G.P., Margison, F.R., O'Dowd, T., Osborn, M. and Moss, S. 1984. The clarification and assessment of a method of psychotherapy. *British Journal of Psychiatry*. **144**(6), pp.567-575.
- Goodman, C.M. 1987. The Delphi technique: A critique. *Journal of Advanced Nursing*. **12**, pp. 729-734.
- Green, D. and Latchford, G. 2012. *Maximising the Benefits of Psychotherapy: A Practice-Based Evidence Approach*. Oxford: Wiley-Blackwell.

- Gupta, U.G. and Clarke, R.E. 1996. Theory and applications of the Delphi technique: A bibliography (1975-1994) *Technological Forecasting and Social Change*. **53**(2), pp.185-211.
- Haddock, G., Beardmore, R., Earnshaw, P., Fitzsimmons, M., Nothard, S., Butler, R., Eisner, E. and Barrowclough, C. 2012. Assessing fidelity to integrated motivational interviewing and CBT therapy for psychosis and substance use: The MI-CBT fidelity scale (MI-CTS). *Journal of Mental Health*. **21**(1), pp.38-48.
- Haddock, G., Devane, S., Bradshaw, T., McGovern, J., Tarrier, N., Kinderman, P., Baguley, I., Lancashire, S. and Harris, N. 2001. An investigation into the psychometric properties of the cognitive therapy scale for psychosis (CTS-Psy). *Behavioural and Cognitive Psychotherapy*. **29**(2), pp.221-233.
- Hallgren, K.A. 2012. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*. **8**(1), pp.23-34.
- Hartley, S., Scarratt, P., Bucci, S., Kelly, J., Mulligan, J., Neil, S.T., Rivers, Z., Taylor, C.D., Welford, M. and Haddock, G. 2014. Assessing therapist adherence to recovery-focused cognitive behavioural therapy for psychosis delivered by telephone with support from a self-help guide: Psychometric evaluations of a New Fidelity Scale. *Behavioural and Cognitive Psychotherapy*. **42**(4), pp.435-451.
- Hasson, D. and Arnetz, B.B. 2005. Validation and findings comparing VAS vs. likert scales for psychosocial measurements. *International Electronic Journal of Health Education*. **8**, pp.178-192.
- Hasson, F., Keeney, S. and McKenna, H. 2000. Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*. **32**(4), pp.1008-1015.
- Hatch-Maillette, M., Burlew, A., Turnbull, S., Robinson, M. and Calsyn, D.A. 2013. Measuring fidelity to a culturally adapted HIV prevention intervention for men in substance abuse treatment. *Journal of Substance Abuse Treatment*. **45**(4), pp.363-369.
- Hegel, M.T., Dietrich, A.J., Seville, J. and Jordan, C. 2004. Training residents in problem-solving treatment of depression: a pilot feasibility and impact study. *Family Medicine*. **36**(3), pp.204-208.
- Henson, R.K. and Roberts, J.K. 2006. Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*. **66**(3), pp.393-416.
- Heppner, P.P., Wampold, B.E., Owen, J., Wang, K.T. and Thompson, M.N. 2016. *Research Design in Counselling (Fourth Edition)*. Boston, MA: Cengage Learning.
- Higgins, J.P.T., Thompson, S.G., Deeks, J.J. and Altman, D.G. 2003. Measuring inconsistency in meta-analyses. *British Medical Journal*. **327**(7414), pp.557-560.
- Hill, C.E., O'Grady, K.E. and Elkin, I. 1992. Applying the collaborative study psychotherapy rating scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. *Journal of Consulting and Clinical Psychology*. **60**(1), pp.73-79.
- Hilsenroth, M.J., Blagys, M.D., Ackerman, S.J., Bonge, D.R. and Blais, M.A. 2005. Measuring psychodynamic-interpersonal and cognitive-behavioral techniques: Development of the comparative psychotherapy process scale. *Psychotherapy: Theory, Research, Practice, Training*. **42**(3), pp.340-356.
- Holey, E.A., Feeley, J.L., Dixon, J. and Whittaker, V.J. 2007. An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC Medical Research Methodology*. [Online]. **7**, pp.1-10. [Accessed 12/09/2018]. Available from: <https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/1471-2288-7-52>
- Hollon, S.D., Evans, M.D., Auerbach, A., DeRubeis, R.J., Elkin, I., Lowery, A., Kriss, M., Grove, W., Tuason, V.B. and Piasecki, J. 1988. *Development of a System for Rating*

- Therapies for Depression: Differentiating Cognitive Therapy, Interpersonal Therapy, and Clinical Management Pharmacotherapy.* Unpublished.
- Horvath, A.O. 2000. The therapeutic relationship: From transference to alliance. *Journal of Clinical Psychology.* **56**(2), pp.163-173.
- Horvath, A.O., Del Re, A.C., Fluckiger, C. and Symonds, D. 2011. Alliance in individual psychotherapy. *Psychotherapy.* **48**(1), pp.9-16.
- Horvath, A.O. and Greenberg, L.S. 1989. Development and validation of the Working Alliance Inventory. *Journal of counseling psychology.* **36**(2), p223.
- Hoyt, M.F., Marmar, C.R., Horowitz, M.J. and Alvarez, W.F. 1981. The Therapist Action Scale: Instruments for the assessment of activities during dynamic psychotherapy. *Psychotherapy: Theory, Research & Practice.* **18**(1), pp.109-116.
- Hsu, C.-C. and Sandford, B.A. 2007. The Delphi technique: making sense of consensus. *Practical Assessment and Evaluation* [Online]. **12**(10), pp.1-8. [Accessed 12/09/2018]. Available from: <http://pareonline.net/pdf/v12n10.pdf>
- Huppert, J.D., Bufka, L.F., Barlow, D.H., Gorman, J.M., Shear, M.K. and Woods, S.W. 2001. Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology.* **69**(5), pp.747-755.
- Hutchings, A. and Raine, R. 2006. A systematic review of factors affecting the judgments produced by formal consensus development methods in health care. *Journal of Health Services Research and Policy.* **11**(3), pp.172-179.
- IBM Corp. 2013. *IBM SPSS Statistics for Windows, Version 22.0.* Armonk, NY: IBM Corp.
- Imel, Z.E., Wampold, B.E., Miller, S.D. and Fleming, R.R. 2008. Distinctions without a difference: Direct comparisons of psychotherapies for alcohol use disorders. *Psychology of Addictive Behaviors.* **22**(4), pp.533-543.
- Imtiaz, S., Shield, K.D., Roerecke, M., Cheng, J., Popova, S., Kurdyak, P., Fischer, B. and Rehm, J. 2016. The burden of disease attributable to cannabis use in Canada in 2012. *Addiction.* **111**(4), pp.653-662.
- Jackson, D.L., Gillaspay, J.A. and Purc-Stephenson, R. 2009. Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods.* **14**(1), pp.6-23.
- Jamieson, S. 2004. Likert scales: How to (ab)use them. *Medical Education.* **38**(12), pp.1217-1218.
- Jisc. 2018. *Online Surveys (Formally Bristol Online Surveys).* [Online]. [Accessed 12/09/2018]. Available from: <https://www.onlinesurveys.ac.uk>
- Joffe, H. 2012. Thematic analysis. In: Harper, D. and Thompson, A. eds. *Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners.* Chichester, UK: Wiley-Blackwell, pp.209-223.
- Joffe, H. and Yardley, L. 2004. Content and thematic analysis. In: Marks, D.F. and Yardley, L. eds. *Research Methods for Clinical and Health Psychology.* London, UK: Sage Publications, pp.56-68.
- Jones, E.E. 2000. *Therapeutic Action: A Guide to Psychoanalytic Therapy.* Northvale, NJ Aronson.
- Jones, E.E. and Pulos, S.M. 1993. Comparing the process in psychodynamic and cognitive-behavioral therapies. *Journal of Consulting and Clinical Psychology.* **61**(2), pp.306-315.
- Kahler, E., Rogausch, A., Brunner, E. and Himmel, W. 2008. A parametric analysis of ordinal quality-of-life data can lead to erroneous results. *Journal of Clinical Epidemiology.* **61**(5), pp.475-480.



- Kalaian, S.A. and Kasim, R.M. 2012. Terminating sequential Delphi survey data collection. *Practical Assessment, Research and Evaluation*. [Online]. **17**(5), pp.1-10. [Accessed 12/09/2018]. Available from: <http://pareonline.net/getvn.asp?v=17&n=5>
- Kao, L.S. and Green, C.E. 2008. Analysis of variance: Is there a difference in means and what does it mean? *The Journal of Surgical Research*. **144**(1), pp.158-170.
- Karno, M.P. and Longabaugh, R. 2007. Does matching matter? Examining matches and mismatches between patient attributes and therapy techniques in alcoholism treatment. *Addiction*. **102**(4), pp.587-596.
- Karterud, S. and Bateman, A. 2010. *Manual for Mentaliseringsbasert Terapi (MBT) og MBT Vurderingsskala: Versjon Individualterapi*. Unpublished.
- Karterud, S., Pedersen, G., Engen, M., Johansen, M.S., Johansson, P.N., Schluter, C., Urnes, O., Wilberg, T. and Bateman, A.W. 2013. The MBT adherence and competence scale (MBT-ACS): Development, structure and reliability. *Psychotherapy Research*. **23**(6), pp.705-717.
- Kazdin, A.E. 2009. Understanding how and why psychotherapy leads to change. *Psychotherapy Research*. **19**(4-5), pp.418-428.
- Kazdin, A.E. and Nock, M.K. 2003. Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry*. **44**(8), pp.1116-1129.
- Kiluk, B.D., Nich, C., Babuscio, T. and Carroll, K.M. 2010. Quality versus quantity: Acquisition of coping skills following computerized cognitive-behavioral therapy for substance use disorders. *Addiction*. **105**(12), pp.2120-2127.
- Kolden, G.G., Klein, M.H., Wang, C. and Austin, S.B. 2011. Congruence/genuineness. *Psychotherapy*. **48**(1), pp.65-71.
- Koo, T.K. and Li, M.Y. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*. **15**(2), pp.155-163.
- Landers, R. 2015. Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *The Winnower*. [Online]. pp.1-4. [Accessed 12/09/2018]. Available from: <http://neoacademic.com/2011/11/16/computing-intraclass-correlations-icc-as-estimates-of-interrater-reliability-in-spss/>
- Lane, C. 2002. *The behaviour change counselling Index (BECCI). Manual for coding behaviour change counselling*. [Online]. [Accessed 12/09/2018]. Available from: <https://www.motivationalinterviewing.org/sites/default/files/BECCIManual.pdf>
- Lane, C., Huws-Thomas, M., Hood, K., Rollnick, S., Edwards, K. and Robling, M. 2005. Measuring adaptations of motivational interviewing: The development and validation of the behavior change counseling index (BECCI). *Patient Education and Counseling*. **56**(2), pp.166-173.
- Lefebvre, C., Manheimer, E., Glanville, J. and on behalf of the Cochrane Information Retrieval Methods Group. 2008. Searching for studies. In: Higgins, J.P.T. and Green, S. eds. *Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series*. [Online]. Chichester: Wiley-Blackwell, pp.95-150. [Accessed 12/09/2018]. Available from: <https://www.radioterapiaitalia.it/wp-content/uploads/2017/01/cochrane-handbook-for-systematic-reviews-of-interventions.pdf>
- Levac, D., Colquhoun, H. and O'Brien, K.K. 2010. Scoping studies: advancing the methodology. *Implementation Science*. [Online]. **5**(1), pp.1-9. [Accessed 12/09/2018]. Available from: <https://doi.org/10.1186/1748-5908-5-69>
- Levy, K.N., Wasserman, R.H., Scott, L.N., Zach, S.E., White, C.N., Cain, N.M., Fischer, C., Carter, C.E., Clarkin, J.F. and Kernberg, O.F. 2006. The development of a measure to assess putative mechanisms of change in the treatment of borderline

- personality disorder. *Journal of the American Psychoanalytic Association*. **54**(4), pp.1325-1331.
- Liese, B.S., Barber, J.P. and Beck, A.T. 1995. *Cognitive Therapy Adherence Competence Scale (CTACS)*. [Online]. [Accessed 12/09/2018]. Available from: <http://bruceliese.com/documents/blpubs/ctac-scale.pdf>
- Liu, J., Tang, W., Chen, G., Lu, Y., Feng, C. and Tu, X.M. 2016. Correlation and agreement: overview and clarification of competing concepts and measures. *Shanghai Archives of Psychiatry*. **28**(2), pp.115-120.
- Longabaugh, R. 2010. Commentary on Kiluk et al. (2010): The emperor has some new clothes – coping is a mediator of cognitive-behavioral therapy's effectiveness. *Addiction*. **105**(12), pp.2128-2129.
- Longabaugh, R., Donovan, D.M., Karno, M.P., McCrady, B.S., Morgenstern, J. and Tonigan, J.S. 2005. Active ingredients: how and why evidence-based alcohol behavioral treatment interventions work. *Alcoholism: Clinical and Experimental Research*. **29**(2), pp.235-247.
- Lu, W., Yanos, P.T., Gottlieb, J.D., Duva, S.M., Silverstein, S.M., Xie, H., Rosenberg, S.D. and Mueser, K.T. 2012. Use of fidelity assessments to train clinicians in the CBT for PTSD program for clients with serious mental illness. *Psychiatric Services*. **63**(8), pp.785-792.
- Ludwig, B. 1997. Predicting the future: Have you considered using the Delphi methodology. *Journal of extension*. [Online]. **35**(5), pp.1-4. [Accessed 12/09/2018]. Available from: [https://www.joe.org/joe/1997october/tt2.php/site\\_urljournal-current-issue.php](https://www.joe.org/joe/1997october/tt2.php/site_urljournal-current-issue.php)
- Lynn, M.R. 1986. Determination and quantification of content validity. *Nursing research*. **35**(6), pp.382-386.
- Lynn, P.R. 2015. Cognitive behavioral theories of depression. In: Wolberg, B. and Stricker, G. eds. *Depressive Disorders: Facts, Theories, and Treatment*. [Online]. Chevy Chase, MD: International Psychotherapy Institute E-Books, pp.1-54. [Accessed 12/09/2018]. Available from: [https://www.israpych.org/books/wp-content/uploads/2015/06/interpersonal\\_psychotherapy\\_for\\_depression\\_-\\_myrna\\_m\\_weissman\\_phd.pdf](https://www.israpych.org/books/wp-content/uploads/2015/06/interpersonal_psychotherapy_for_depression_-_myrna_m_weissman_phd.pdf)
- Madson, M.B. and Campbell, T.C. 2006. Measures of fidelity in motivational enhancement: A systematic review. *Journal of Substance Abuse Treatment*. **31**(1), pp.67-73.
- Madson, M.B., Campbell, T.C., Barrett, D.E., Brondino, M.J. and Melchert, T.P. 2005. Development of the Motivational Interviewing Supervision and Training Scale. *Psychology of Addictive Behaviors*. **19**(3), pp.303-310.
- Madson, M.B. and Loignon, A.C. 2007. *Rater Manual for the Motivational Interviewing Supervision and Training Scale – Revised*. Unpublished.
- Magill, M., Kiluk, B.D., McCrady, B., Tonigan, J.S. and Longabaugh, R. 2015. Active ingredients of treatment and client mechanisms of change in behavioral treatments for alcohol use disorders: Progress 10 years later. *Alcoholism, Clinical and Experimental Research*. **39**(10), pp.1852-1862.
- Mandrekar, J.N. 2011. Measures of interrater agreement. *Journal of Thoracic Oncology*. **6**(1), pp.6-7.
- Martino, S., Ball, S., Nich, C., Frankforter, T.L. and Carroll, K.M. 2009a. Correspondence of motivational enhancement treatment integrity ratings among therapists, supervisors, and observers. *Psychotherapy Research*. **19**(2), pp.181-193.
- Martino, S., Ball, S.A., Nich, C., Frankforter, T.L. and Carroll, K.M. 2008. Community program therapist adherence and competence in motivational enhancement therapy. *Drug and Alcohol Dependence*. **96**(1-2), pp.37-48.
- Martino, S., Paris, M., Añez, L., Canning-Ball, M. and Carroll, K.M. 2009b. *Independent Rater Guide: Manual for Rating Clinician Adherence and Competence*. Unpublished.

- Mays, N. and Pope, C. 2000. Qualitative research in health care: Assessing quality in qualitative research. *BMJ*. **320**, pp.50-52.
- Mays, N. and Pope, C. 2006. Quality in qualitative health research. In: Pope, C. and Mays, N. eds. *Qualitative Research in Health Care (Third Edition)*. Oxford, UK: Blackwell Publishing, pp.82-101.
- McDonagh, M., Peterson, K., Raina, P., Chang, S. and Shekelle, P. 2013. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews: Avoiding Bias in Selecting Studies*. [Online]. Rockville, MD: Agency for Healthcare Research and Quality. [Accessed 12/09/2018]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK126701/>
- McGraw, K. and Wong, S.P. 1996. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. **1**(1), pp.30-46.
- McGuire, A.B., Stull, L.G., Mueser, K.T., Santos, M., Mook, A., Rose, N., Tunze, C., White, L.M. and Salyers, M.P. 2012. Development and reliability of a measure of clinician competence in providing illness management and recovery. *Psychiatric Services*. **63**(8), pp.772-778.
- McHugh, M.L. 2012. Interrater reliability: The kappa statistic. *Biochemia Medica*. **22**(3), pp.276-282.
- McKenna, H.P. 1994. The Delphi technique: a worthwhile research approach for nursing? *Journal of Advanced Nursing*. **19**(6), pp.1221-1225.
- Mee-Lee, D., McLellan, A.T. and Miller, S.D. 2010. What works in substance abuse and dependence treatment. In: Duncan, B.L., et al. eds. *The Heart and Soul of Change Second Edition: Delivering What Works in Therapy*. London: American Psychological Association, pp.393-417.
- Meline, T. 2006. Selecting studies for systematic review: Inclusion and exclusion criteria. *Contemporary Issues in Communication Science and Disorders*. **33**, pp.21-27.
- Mercer, D., Calvo, N., Krakauer, I.D. and Barber, J.P. 1995. *Adherence/Competence Scale for Individual Drug Counseling (IDC) for Cocaine Dependence*. Unpublished.
- Messick, S. 1995. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*. **50**(9), pp.741-749.
- Michie, S., Abraham, C., Eccles, M.P., Francis, J.J., Hardeman, W. and Johnston, M. 2011. Strengthening evaluation and implementation by specifying components of behaviour change interventions: a study protocol. *Implementation Science* [Online]. **6**(10), pp.1-8. [Accessed 12/09/2018]. Available from: <https://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-6-10>
- Middel, B. and van Sonderen, E. 2002. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care*. [Online]. **2**, pp.1-18. [Accessed 12/09/2018]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1480399/pdf/ijic2002-200215.pdf>
- Middleton, W., Tober, G., Frier, E. and Finnegan, O., with the UKATT Process Data Sub-Group. 2001. *UKATT Process Rating Manual*. Unpublished.
- Miller, W.R., Moyers, T.B., Ernst, D. and Amrhein, P. 2008a. *Manual for the Motivational Interviewing Skill Code (MISC) Version 2.1*. [Online]. [Accessed 12/09/2018]. Available from: <http://casaa.unm.edu/download/misc.pdf>
- Miller, W.R., Moyers, T.B., Manuel, J.K., Christopher, P. and Amrhein, P. 2008b. *Revision for Client Language Coding: MISC 2.1; Client Language Assessment in Motivational Interviewing (CLAMI) Segment* [Online]. [Accessed 12/08/2018]. Available from: <https://casaa.unm.edu/download/clami.pdf>

- Miller, W.R. and Rollnick, S. 2013. *Motivational Interviewing: Helping People Change (Third Edition)*. London: The Guilford Press.
- Miller, W.R., Zweben, A., DiClemente, C.C. and Rychtarik, R. 1992. *Motivational Enhancement Therapy Manual: A Clinical Research Guide for Therapists Treating Individuals with Alcohol Abuse and Dependence. Project MATCH Monograph Series, Volume 2*. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.
- Milne, D. 2007. An empirical definition of clinical supervision. *British Journal of Clinical Psychology*. **46**(Pt 4), pp.437-447.
- Mircioiu, C. and Atkinson, J. 2017. A comparison of parametric and non-parametric methods applied to a Likert scale. *Pharmacy: Journal of Pharmacy, Education and Practice*. [Online]. **5**(2), pp.1-12. [Accessed 12/09/2018]. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5597151/>
- Moorey, S. 2002. Cognitive therapy. In: Dryden, W. ed. *Handbook of Individual Therapy (Fourth Edition)*. London: Sage Publications.
- Moos, R.H. 2007. Theory-based active ingredients of effective treatments for substance use disorders. *Drug and Alcohol Dependence*. **88**(2-3), pp.109-121.
- Morgenstern, J. and Longabaugh, R. 2000. Cognitive-behavioral treatment for alcohol dependence: A review of evidence for its hypothesized mechanisms of action. *Addiction*. **95**(10), pp.1475-1490.
- Morse, J.M. 2003. Principles of mixed methods and multimethod research design. In: Tashakkori, A. and Teddie, C. eds. *Handbook of Mixed Methods in Social and Behavioral Research*. London: Sage, pp.189-208.
- Moyers, T., Martin, T., Catley, D., Harris, K.J. and Ahluwalia, J.S. 2003. Assessing the integrity of motivational interviewing interventions: Reliability of the motivational interviewing skills code. *Behavioural and Cognitive Psychotherapy*. **31**(2), pp.177-184.
- Moyers, T.B. 2004. *Global Rating of Motivational Interviewing Therapist (GROMIT)*. [Online]. [Accessed 12/09/2018]. Available from: <http://casaa.unm.edu/download/gromit.pdf>
- Moyers, T.B., Manuel, J.K. and Ernst, D. 2015. *Motivational Interviewing Treatment Integrity Coding Manual 4.2.1*. Unpublished.
- Moyers, T.B., Martin, T., Manuel, J.K., Hendrickson, S.M.L. and Miller, W.R. 2005. Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment*. **28**(1), pp.19-26.
- Murphy, M.K., Black, N.A., Lamping, D.L., McKee, C.M., Sanderson, C.F., Askham, J. and Marteau, T. 1998. Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*. **2**(3), pp.1-88.
- National Institute for Health and Clinical Excellence. 2011. *Alcohol-use disorders: Diagnosis, assessment and management of harmful drinking and alcohol dependence [CG115]* [Online]. London: National Institute for Health and Care Excellence. [Accessed 23/07/2018]. Available from: <https://www.nice.org.uk/guidance/cg115/chapter/1-Guidance#interventions-for-alcohol-misuse>
- Neenan, M. and Dryden, W. 2006. *Cognitive Therapy in a Nutshell*. London: SAGE.
- Nestor, P.G. and Schutt, R.K. 2015. *Research Methods in Psychology: Investigating Human Behavior (Second Edition)*. California: SAGE publications.
- NHS Digital. 2018. *Statistics on Alcohol: England*. [Online]. [Accessed 12/09/2018]. Available from: <https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-alcohol/2018>
- NIST/SEMATECH. 2013. *e-Handbook of Statistical Methods*. [Online]. [Accessed 12/09/2018]. Available from: <https://www.itl.nist.gov/div898/handbook/index.htm>

- Noordzij, M., Tripepi, G., Dekker, F.W., Zoccali, C., Tanck, M.W. and Jager, K.J. 2010. Sample size calculations: basic principles and common pitfalls. *Nephrology Dialysis Transplantation*. **25**(5), pp.1388-1393.
- Norcross, J.C. 2010. The therapeutic relationship. In: Duncan, B.L., et al. eds. *The Heart and Soul of Change: Delivering What Works in Therapy (Second Edition)*. Washington, DC: American Psychological Association, pp.113-141.
- Norcross, J.C. and Wampold, B.E. 2011. What works for whom: Tailoring psychotherapy to the person. *Journal of Clinical Psychology*. **67**(2), pp.127-132.
- Norman, G. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*. **15**(5), pp.625-632.
- Nowell, L.S., Norris, J.M., White, D.E. and Moules, N.J. 2017. Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*. **16**(1), pp.1-13.
- Nuro, K.F., Maccarelli, L., Martino, S., Ball, S., Baker, S.M., Rounsaville, B.J. and Carroll, K.M. 2005. *Yale Adherence and Competence Scale (YACSII) Guidelines Second Edition*. [Online]. [Accessed 12/09/2018]. Available from: [http://mirecc.va.gov/MIRECC/visn1/docs/products/Yale\\_Adherence\\_and\\_Competence\\_Scale\\_II\\_Guidelines.pdf](http://mirecc.va.gov/MIRECC/visn1/docs/products/Yale_Adherence_and_Competence_Scale_II_Guidelines.pdf)
- O'Malley, S.S., Foley, S.H., Rounsaville, B.J., Watkins, J.T., Sotsky, S.M., Imber, S.D. and Elkin, I. 1988. Therapist competence and patient outcome in interpersonal psychotherapy of depression. *Journal of Consulting and Clinical Psychology*. **56**(4), pp.496-501.
- O'Rourke, N. and Hatcher, L. 2013. *A Step-by Step Approach to Using SAS for Factor Analysis and Structural Equation Modelling*. Cary, NC: SAS Institute Inc.
- Onwumere, J., Kuipers, E., Gamble, C., Jolley, S., Smith, B., Rollinson, R., Steel, C., Fowler, D., Bebbington, P., Dunn, G., Freeman, D. and Garety, P. 2009. Family interventions in psychosis: A scale to measure therapist adherence. *Journal of Family Therapy*. **31**(3), pp.270-283.
- Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N. and Hoagwood, K. 2015. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health*. **42**(5), pp.533-544.
- Pantalon, M.V., Martino, S., Dziura, J., Li, F.Y., Owens, P.H., Fiellin, D.A., O'Connor, P.G. and D'Onofrio, G. 2012. Development of a scale to measure practitioner adherence to a brief intervention in the emergency department. *Journal of Substance Abuse Treatment*. **43**(4), pp.382-388.
- Patton, M.Q. 2002. *Qualitative Research and Evaluation Methods (Third Edition)*. Thousand Oaks, California: Sage Publications.
- Perepletchikova, F. and Kazdin, A.E. 2005. Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice*. **12**(4), pp.365-383.
- Perepletchikova, F., Treat, T.A. and Kazdin, A.E. 2007. Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*. **75**(6), pp.829-841.
- Petrik, A.M., Kazantzis, N. and Hofmann, S.G. 2013. Distinguishing integrative from eclectic practice in cognitive behavioral therapies. *Psychotherapy*. **50**(3), pp.392-397.
- Petry, N.M. and Stitzer, M.L. 2002. *Contingency Management: Using Motivational Incentives to Improve Drug Abuse Treatment*. West Haven, CT: Yale University Psychotherapy Development Center.
- Pilgrim, D. 2014. Some implications of critical realism for mental health research. *Social Theory and Health*. **12**(1), pp.1-21.

- Powell, C. 2003. The Delphi technique: Myths and realities. *Journal of Advanced Nursing*. **41**(4), pp.376-382.
- Project MATCH Research Group. 1997. Matching Alcoholism Treatments to Client Heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol*. **58**(1), pp.7-29.
- Prowse, P.-T. and Nagel, T. 2014. Developing an instrument for assessing fidelity of motivational care planning: The aboriginal and islander mental health initiative adherence scale. *International Journal of Mental Health Systems*. **8**, p36.
- Prowse, P.T. and Nagel, T. 2013. *Motivational Care Planning -Adherence Scale Protocol Manual*. Unpublished.
- Raistrick, D., Heather, N. and Godfrey, C. 2006. *Review of the effectiveness of treatment for alcohol problems*. [Online]. [Accessed 12/09/2018]. Available from: <http://www.dldocs.stir.ac.uk/documents/alcoeffective.pdf>
- Reid, N. 1993. The Delphi technique: Its contribution to the evaluation of professional practice. In: Ellis, R. ed. *Professional Competence and Quality Assurance in the Caring Professions*. London: Chapman and Hall, pp.230-254.
- Rosengren, D.B., Baer, J.S., Hartzler, B., Dunn, C.W., Wells, E.A. and Ogle, R. 2009. *Video Assessment of Simulated Encounters (VASE-R) Administration and Scoring Manual*. [Online]. [Accessed 12/09/2018]. Available from: [http://adai.washington.edu/instruments/PDF/VASERScoreManual\\_145.pdf](http://adai.washington.edu/instruments/PDF/VASERScoreManual_145.pdf)
- Rosengren, D.B., Hartzler, B., Baer, J.S., Wells, E.A. and Dunn, C.W. 2008. The video assessment of simulated encounters-revised (VASE-R): Reliability and validity of a revised measure of motivational interviewing skills. *Drug and Alcohol Dependence*. **97**(1-2), pp.130-138.
- Rowe, G. and Wright, G. 1999. The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*. **15**(4), pp.353–375.
- Rubio, D.M., Berg-Weger, M., Tebb, S.S., Lee, E.S. and Rauch, S. 2003. Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*. **27**(2), pp.94-104.
- Schmucker, C., Bluemle, A., Briel, M., Portalupi, S., Lang, B., Motschall, E., Schwarzer, G., Bassler, D., Mueller, K.F., von Elm, E. and Meerpohl, J.J. 2013. A protocol for a systematic review on the impact of unpublished studies and studies published in the gray literature in meta-analyses. *Systematic Reviews*. **2**, pp.24-24.
- Schoenwald, S.K., Garland, A.F., Chapman, J.E., Frazier, S.L., Sheidow, A.J. and Southam-Gerow, M.A. 2011. Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*. **38**(1), pp.32-43.
- Schreiber, J.B., Nora, A., Stage, F.K., Barlow, E.A. and King, J. 2006. Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*. **99**(6), pp.323-338.
- Segal, Z.V., Teasdale, J.D., Williams, J.M. and Gemar, M.C. 2002. The mindfulness-based cognitive therapy adherence scale: Inter-rater reliability, adherence to protocol and treatment distinctiveness. *Clinical Psychology and Psychotherapy*. **9**(2), pp.131-138.
- Shapiro, D.A. and Startup, M. 1990. *Raters' Manual for the Sheffield Psychotherapy Rating Scale (SPRS)*. Sheffield, UK: MRC/ESRC Social & Applied Psychology Unit, University of Sheffield.
- Shapiro, D.A. and Startup, M. 1992. Measuring therapist adherence in exploratory psychotherapy. *Psychotherapy Research*. **2**(3), pp.193-203.
- Shedler, J. 2010. The efficacy of psychodynamic psychotherapy. *American Psychologist*. **65**(2), pp.98-109.

- Shrout, P.E. and Fleiss, J.L. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*. **86**(2), pp.420-428.
- Song, M.-K., Happ, M.B. and Sandelowski, M. 2010. Development of a tool to assess fidelity to a psycho-educational intervention. *Journal of Advanced Nursing*. **66**(3), pp.673-682.
- SPR Project Staff. 1984. *Rating Manual for the Collaborative Study Psychotherapy Rating Scale - Form 6 (CSPRS-6)*. Unpublished.
- Startup, M., Jackson, M. and Pearce, E. 2002. Assessing therapist adherence to cognitive-behaviour therapy for psychosis. *Behavioural and Cognitive Psychotherapy*. **30**(3), pp.329-339.
- Startup, M. and Shapiro, D.A. 1993. Therapist treatment fidelity in prescriptive vs. exploratory psychotherapy. *British Journal of Clinical Psychology*. **32**(4), pp.443-456.
- StataCorp. 2013a. *Kappa - Interrater Agreement*. [Online]. [Accessed 12/09/2018]. Available from: <https://www.stata.com/manuals13/rkappa.pdf>
- StataCorp. 2013b. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.
- Stein, K.F., Sargent, J.T. and Rafaels, N. 2007. Intervention research: establishing fidelity of the independent variable in nursing clinical trials. *Nursing Research*. **56**(1), pp.54-62.
- Strang, J. and McCambridge, J. 2004. Can the practitioner correctly predict outcome in motivational interviewing? *Journal of Substance Abuse Treatment*. **27**(1), pp.83-88.
- Strauss, M.E. and Smith, G.T. 2009. Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*. **5**, pp.1-25.
- Streiner, D.L., Norman, G.R. and Cairney, J. 2015. *Health Measurement Scales: A Practical Guide to their Development and Use (Fifth Edition)*. Oxford: Oxford University Press.
- Suri, H. 2011. Purposeful sampling in qualitative research synthesis. *Qualitative Research Journal*. **11**(2), pp.63-75.
- Svartberg, M. 1989. Manualization and competence monitoring of short-term anxiety-provoking psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*. **26**(4), pp.564-571.
- Tadic, M. and Despland, J.N. 2001. *Manuel de compétence pour l'investigation psychodynamique brève*. . Unpublished.
- Tadic, M., Drapeau, M., Solai, S., De Roten, Y. and Despland, J.N. 2003. Development of a competence scale for brief psychodynamic investigation: A pilot study. *Schweizer Archiv fur Neurologie und Psychiatrie*. **154**(1), pp.28-36.
- Taleb, N.M. 2010. *The Black Swan: The Impact of the Highly Improbable*. London, England: Penguin Books.
- Tang, W., Hu, J., Zhang, H., Wu, P. and He, H. 2015. Kappa coefficient: A popular measure of rater agreement. *Shanghai Archives of Psychiatry*. **27**(1), pp.62-67.
- Tavakol, M. and Dennick, R. 2011. Making sense of Cronbach's alpha. *International Journal of Medical Education*. **2**, pp.53-55.
- Tober, G., Clyne, W., Finnegan, O., Farrin, A. and in collaboration with the UKATT Research Team. 2008. Validation of a scale for rating the delivery of psycho-social treatments for alcohol dependence and misuse: The UKATT Process Rating Scale (PRS). *Alcohol and Alcoholism*. **43**(6), pp.675-682.
- Tober, G. and Crosby, H.F. 2011. *AESOPS Process Rating Manual Version 5*. Unpublished.
- Tober, G. and Crosby, H.F. 2014. *ADAPTA Process Rating Scale Manual*. Unpublished.
- Torrey, A.R. 2011. *Development and Psychometric Evaluation of a Structured Instrument to Assess the Treatment Fidelity of a Brief Opportunistic Intervention Designed to Reduce Substance Use Among Pregnant Women*. In partial fulfillment of the requirements for the degree of Doctor of Philosophy thesis, Duquesne University.

- Torrey, A.R. 2012. Treatment fidelity instrument to measure a brief opportunistic intervention for prenatal substance use. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*. **41**(6), pp.809-820.
- Tracey, T.J. and Kokotovic, A.M. 1989. Factor structure of the Working Alliance Inventory. *Psychological Assessment*. **3**(1), pp.207-210.
- Trevethan, R. 2017. Intraclass correlation coefficients: clearing the air, extending some cautions, and making some requests. *Health Services and Outcomes Research Methodology*. **17**(2), pp.127-143.
- Tryon, G.S. and Winograd, G. 2011. Goal consensus and collaboration. In: Norcross, J.C. ed. *Psychotherapy Relationships that Work: Evidence-Based Responsiveness (Second Edition)*. New York: Oxford University Press, pp.153-167.
- Tweed, A., Graber, R. and Wang, M. 2010. Assessing trainee clinical psychologists' clinical competence. *Psychology Learning & Teaching*. **9**(2), pp.50-60.
- Tyrer, P., Jones, V., Thompson, S., Catalan, J., Schmidt, U., Davidson, K., Knapp, M. and Ukoumunne, O., (on behalf of the POPMACT group). 2003. Service validation in baseline variables and prediction of risk in a randomised controlled trial of psychological treatment in repeated parasuicide: The POPMACT study. *International Journal of Social Psychiatry*. **49**(1), pp.58-69.
- UKATT Research Team. 2001. United Kingdom alcohol treatment trial (UKATT): Hypotheses, design and methods. *Alcohol & Alcoholism*. **36**(1), pp.11-21.
- UKATT Research Team. 2005. Effectiveness of treatment for alcohol problems: findings of the randomised UK alcohol treatment trial (UKATT). *British Medical Journal*. [Online]. **331**(7516), pp.1-5. [Accessed 12/09/2018]. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1200586/>
- Volkow, N.D., Baler, R.D., Compton, W.M. and Weiss, S.R.B. 2014. Adverse health effects of marijuana use. *New England Journal of Medicine*. **370**(23), pp.2219-2227.
- von Consbruch, K., Clark, D.M. and Stangier, U. 2012. Assessing therapeutic competence in cognitive therapy for social phobia: Psychometric properties of the cognitive therapy competence scale for social phobia (CTCS-SP). *Behavioural and Cognitive Psychotherapy*. **40**(2), pp.149-161.
- von der Gracht, H.A. 2012. Consensus measurement in Delphi studies: Review and implications for future quality assurance. *Technological Forecasting and Social Change*. **79**(8), pp.1525-1536.
- Walters, S.J. 2009. Consultants' forum: Should post hoc sample size calculations be done? *Pharmaceutical Statistics*. **8**, pp.163-169.
- Waltz, J., Addis, M.E., Koerner, K. and Jacobson, N.S. 1993. Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*. **61**(4), pp.620-630.
- Wampold, B.E. 2015. How important are the common factors in psychotherapy? An update. *World Psychiatry*. **14**(3), pp.270-277.
- Wampold, B.E. and Imel, Z.E. 2015. *The Great Psychotherapy Debate: The Evidence for What Makes Psychotherapy Work*. Second Edition ed. New York: Routledge.
- Wampold, B.E., Mondin, G.W., Moody, M., Stich, F., Benson, K. and A, H. 1997. A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes". *Psychological Bulletin*. **122**(3), pp.203-215.
- Warrens, M.J. 2012. Cohen's linearly weighted kappa is a weighted average. *Advances in Data Analysis and Classification*. **6**(1), pp.67-79.
- Watson, J., Crosby, H., Dale, V., Tober, G., Wu, Q., Lang, J., McGovern, R., Newbury-Birch, D., Parrott, S. and Bland, J. 2013a. AESOPS: a randomised controlled trial of the clinical effectiveness and cost-effectiveness of opportunistic screening and stepped care interventions for older hazardous alcohol users in primary care. *Health Technology Assessment*. **17**(25), pp.1-158.



- Watson, J., Tober, G., Raistrick, D., Mdege, N., Dale, V., Crosby, H., Godfrey, C., Lloyd, C., Toner, P. and Parrott, S. 2013b. An alcohol-focused intervention versus a healthy living intervention for problem drinkers identified in a general hospital setting (ADAPTA): study protocol for a randomized, controlled pilot trial. *Trials*. [Online]. **14**(1), p117. [Accessed 12/09/2018]. Available from: <https://trialsjournal.biomedcentral.com/track/pdf/10.1186/1745-6215-14-117?site=trialsjournal.biomedcentral.com>
- Watson, J.M., Fairhurst, C., Li, J., Tober, G., Crosby, H., Lloyd, C., Godfrey, C., Mdege, N.D., Dale, V., Toner, P., Parrott, S. and Raistrick, D. 2015. ADAPTA: A pilot randomised controlled trial of an alcohol-focused intervention versus a healthy living intervention for problem drinkers identified in a general hospital setting. *Drug and Alcohol Dependence*. **154**, pp.117-124.
- Watson, P.F. and Petrie, A. 2010. Method agreement analysis: A review of correct methodology. *Theriogenology*. **73**, pp.1167–1179.
- Webb, N.M., Shavelson, R.J. and Haertel, E.H. 2006. Reliability coefficients and generalizability theory. In: Rao, C.R. and Sinharay, S. eds. *Handbook of Statistics*. [Online]. Elsevier, pp.81-124. [Accessed 12/09/2018]. Available from: <http://www.sciencedirect.com/science/article/pii/S0169716106260048>
- Weir, J.P. 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*. **19**(1), pp.231-240.
- Weissman, M.M. and Klerman, G.L. 2015. Interpersonal psychotherapy for depression. In: Wolberg, B. and Stricker, G. eds. *Depressive Disorders: Facts, Theories, and Treatment*. [Online]. Chevy Chase, MD: International Psychotherapy Institute E-Books, pp.1-36. [Accessed 12/09/2018]. Available from: [https://www.israppsych.org/books/wp-content/uploads/2015/06/interpersonal\\_psychotherapy\\_for\\_depression\\_-\\_myrna\\_m\\_weissman\\_phd.pdf](https://www.israppsych.org/books/wp-content/uploads/2015/06/interpersonal_psychotherapy_for_depression_-_myrna_m_weissman_phd.pdf)
- Wetzal, A.P. 2012. Factor analysis methods and validity evidence: A review of instrument development across the medical education continuum. *Academic Medicine*. **87**(8), pp.1060-1069.
- Willig, C. 2013. *Introducing Qualitative Research in Psychology*. Third Edition ed. Maidenhead: McGraw-Hill Education.
- World Health Organization. 2011. *Global Status Report on Alcohol and Health*. [Online]. [Accessed 12/09/2018]. Available from: [http://www.who.int/substance\\_abuse/publications/global\\_alcohol\\_report/msbgsr\\_uprofiles.pdf](http://www.who.int/substance_abuse/publications/global_alcohol_report/msbgsr_uprofiles.pdf)
- Wright, D.B., London, K. and Field, A.P. 2011. Using bootstrap estimation and the plug-in principle for clinical psychology data. *Journal of Experimental Psychopathology*. **2**(2), pp.252-270.
- Yardley, L. and Marks, D.F. 2004. Introduction to research methods in clinical and health psychology. In: Marks, D.F. and Yardley, L. eds. *Research Methods for Clinical and Health Psychology*. London, UK: Sage Publications, pp.1-20.
- Young, J. and Beck, A.T. 1980. Cognitive Therapy Scale: Rating Manual. *Unpublished manuscript*.

## **Appendix A**

### **Chapter 2 Appendices**

Appendix A focuses on the appendices for chapter 2, the research overview. The regulatory approvals obtained for the project are presented. Approvals were given from:

- i) Yorkshire and the Humber (Leeds West) National Research Ethics Service (NRES) Committee for the project (NRES committee) (p.193).
- ii) Leeds and York Partnership NHS Foundation Trust Research and Development (R&D) Department for the project (NHS permission) (p.198).
- iii) School of Medicine Research Ethics Committee at the University of Leeds for Study 3, the Delphi exercise (University ethics) (p.202).
- iv) Yorkshire and the Humber (Leeds West) NRES committee for recruitment of clients' friends and family members in Study 4, testing reliability and validity (Amendment 1) (204).
- v) Yorkshire and the Humber (Leeds West) NRES committee for recruitment at a non-NHS drug and alcohol service in Wales (Amendment 2) (p.206).

## A.1 Regulatory approvals – NRES committee



**Health Research Authority**  
NRES Committee Yorkshire & The Humber - Leeds West  
Room 001, Jarrow Business Centre  
Rolling Mill Road  
Jarrow  
Tyne and Wear  
NE32 3DT

Telephone: 0191 4283548

17 March 2015

Miss Helen F. Crosby  
Researcher/PhD Student  
University of Leeds  
Leeds Institute of Health Sciences  
Room G.02, Charles Thackrah Building  
101 Clarendon Road  
Leeds  
LS2 9NG

Dear Miss Crosby

**Study title:** The development of the Addiction Therapist Rating Scale (ATRS): a rating scale for evaluating the delivery of evidence based treatments used in routine practice for alcohol and illicit drug use problems.

**REC reference:** 15/YH/0037

**IRAS project ID:** 165130

Thank you for your letter of 9<sup>th</sup> March, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information was considered in correspondence by a Sub-Committee of the REC. A list of the Sub-Committee members is attached.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager, Miss Sarah Grimshaw, [nrescommittee.yorkandhumber-leedswest@nhs.net](mailto:nrescommittee.yorkandhumber-leedswest@nhs.net). Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

### Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation,

subject to the conditions specified below.

#### Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

*Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.*

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at <http://www.rforum.nhs.uk>.

*Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.*

*For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.*

*Sponsors are not required to notify the Committee of approvals from host organisations*

#### Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact [hra.studyregistration@nhs.net](mailto:hra.studyregistration@nhs.net). The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from NRES. Guidance on where to register is provided on the HRA website.

**It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).**

#### Ethical review of research sites

NHS sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

#### Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	Version	Date
Covering letter on headed paper [Covering letter to REC outlining requested changes]	1	27 February 2015
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [UoL Confirmation of Liability Insurance]		17 September 2014
Letters of invitation to participant [Study 2a Treatment Session Recordings Service User Cover Letter]	1	25 February 2015
Non-validated questionnaire [Study 2a Treatment Session Recordings Therapist Questionnaire]	1	22 September 2014
Non-validated questionnaire [Study 3 Stage 1 Staff Questionnaire]	1	22 September 2014
Non-validated questionnaire [Study 3 Stage 2 Staff Questionnaire]	1	22 September 2014
Non-validated questionnaire [Study 3 Stage 2 Contact Information Form]	1	22 September 2014
Non-validated questionnaire [Study 2a Treatment Session Recordings Service User Questionnaire]	1	22 September 2014
Other [Bewick CV]		07 January 2015
Other [Email from the REC re PALS equivalent]	1	02 March 2015
Participant consent form [Study 2a Treatment Session Recordings Service User Consent Form]	1	22 September 2014
Participant consent form [Study 3 Stage 1 Service User Consent Form]	1	22 September 2014
Participant consent form [Study 1 Item Piloting Consent Form]	2	25 February 2015
Participant consent form [Study 2a Treatment Session Recordings Therapist Consent Form]	2	25 February 2015
Participant consent form [Study 2b Face Validity Consent Form]	2	25 February 2015
Participant consent form [Study 2c Concurrent Validity Consent Form]	2	25 February 2015
Participant consent form [Study 3 Stage 1 Therapist Consent Form]	2	25 February 2015
Participant consent form [Study 3 Stage 2 Consent Form]	2	25 February 2015
Participant information sheet (PIS) [Study 2a Treatment Session Recordings Service User Information Sheet]	1	22 September 2014
Participant information sheet (PIS) [Study 3 Stage 1 Service User Information Sheet]	1	22 September 2014
Participant information sheet (PIS) [Study 1 Item Piloting Information Sheet]	2	25 February 2015
Participant information sheet (PIS) [Study 2a Treatment Session Recordings Service User Information Sheet]	1	22 September 2014
Participant information sheet (PIS) [Study 2a Treatment Session Recordings Therapist Information Sheet]	2	25 February 2015
Participant information sheet (PIS) [Study 2b Face Validity Information Sheet]	2	25 February 2015
Participant information sheet (PIS) [Study 2c Concurrent Validity Information Sheet]	2	25 February 2015

Participant information sheet (PIS) [Study 3 Stage 1 Therapist Information Sheet]	2	25 February 2015
Participant information sheet (PIS) [Study 3 Stage 2 Information Sheet]	2	25 February 2015
REC Application Form [REC_Form_13012015]		13 January 2015
Research protocol or project proposal [The ATRS Project Proposal]	2	25 February 2015
Summary CV for Chief Investigator (CI) [CV for CI]		07 January 2015
Summary CV for supervisor (student research) [CV for Lead Supervisor]		07 January 2015

#### Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

#### After ethical review

##### Reporting requirements

The attached document "*After ethical review – guidance for researchers*" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

#### User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:

<http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/>

#### HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at <http://www.hra.nhs.uk/hra-training/>

15/YH/0037

Please quote this number on all correspondence

With the Committee's best wishes for the success of this project.

Yours sincerely



pp

**Dr Vera Neumann**  
Vice Chair

Email: [nrescommittee.yorkandhumber-leedswest@nhs.net](mailto:nrescommittee.yorkandhumber-leedswest@nhs.net)

*Enclosures: List of names and professions of members  
who were present at the meeting and those who submitted written  
comments*

*"After ethical review – guidance for  
researchers"*

*Copy to: Clare Skinner, University of Leeds  
Ms Sinead Audsley, Leeds and York Partnership NHS Foundation Trust*

## A.2 Regulatory approvals – NHS permission



Leeds and York Partnership **NHS**  
NHS Foundation Trust

Our Ref: 2015/553/P

Research & Development  
St Mary's House  
St Mary's Road  
Leeds  
LS7 3JX

E-mail: [sinead.audsley@nhs.net](mailto:sinead.audsley@nhs.net)  
Direct Line: 0113 295 2387  
FAX: 0113 2954466

25/03/2015

Miss Helen F. Crosby  
Research Fellow  
Leeds Institute of Health Sciences  
University of Leeds  
Room G.02, Charles Thackrah Building  
101 Clarendon Road  
Leeds  
LS2 9NG

Dear Miss Crosby

**Re: The development of the Addiction Therapist Rating Scale (ATRS): a rating scale for evaluating the delivery of evidence based treatments used in routine practice for alcohol and illicit drug use problems.**

UKCRN No: 18297

REC Reference: 15/YH/0037

IRAS/CSP Reference: 165130

Following the recent review of the above project I am pleased to inform you that the above project complies with Research Governance standards, and NHS Permission has been granted on behalf of Trust management. We now have all the relevant documentation relating to the above project. As such your project may now begin within Leeds and York Partnership NHS Foundation Trust (LYPFT).

Accruals officers: please ensure all accruals for this site are linked to LYPFT with NIHR site identifier N0007895. Such accruals to the study should be reported to both NIHR and this department on a monthly basis.

It is anticipated that first reported accrual will occur within 30 days of the date of this letter, in line with national targets for recruitment to NIHR Portfolio projects.

The final list of documents reviewed and approved is as follows:

Document	Version	Date
REC approval letter		17 March 2015
Research protocol	2	25 February 2015
Evidence of Sponsor insurance		17 September 2014



Study 2a Treatment Session Recordings Service User Cover Letter	1	25 February 2015
Study 2a Treatment Session Recordings Therapist Questionnaire	1	22 September 2014
Study 3 Stage 1 Staff Questionnaire	1	22 September 2014
Study 3 Stage 2 Staff Questionnaire	1	22 September 2014
Study 3 Stage 2 Contact Information Form	1	22 September 2014
Study 2a Treatment Session Recordings Service User Questionnaire	1	22 September 2014
Study 2a Treatment Session Recordings Service User Consent Form	1	22 September 2014
Study 3 Stage 1 Service User Consent Form	1	22 September 2014
Study 1 Item Piloting Consent Form	2	25 February 2015
Study 2a Treatment Session Recordings Therapist Consent Form	2	25 February 2015
Study 2b Face Validity Consent Form	2	25 February 2015
Study 2c Concurrent Validity Consent Form	2	25 February 2015
Study 3 Stage 1 Therapist Consent Form	2	25 February 2015
Study 3 Stage 2 Consent Form	2	25 February 2015
Study 2a Treatment Session Recordings Service User Information Sheet	1	22 September 2014
Study 3 Stage 1 Service User Information Sheet	1	22 September 2014
Study 1 Item Piloting Information Sheet	2	25 February 2015
Study 2a Treatment Session Recordings Therapist Information Sheet	2	25 February 2015
Study 2b Face Validity Information Sheet	2	25 February 2015
Study 2c Concurrent Validity Information Sheet	2	25 February 2015
Study 3 Stage 1 Therapist Information Sheet	2	25 February 2015
Study 3 Stage 2 Information Sheet	2	25 February 2015
Summary CV for Chief Investigator		07 January 2015
Summary CV for supervisor		07 January 2015
Bewick CV		07 January 2015

This approval is granted subject to the following conditions:

- You must comply with the terms of your ethical approval (where applicable). Failure to do this will lead to permission to carry out this project being withdrawn. If you make any substantive changes to your protocol you must inform the relevant ethics committee and us immediately.
- You must comply with the Trust's procedures on project monitoring and audit.
- You must comply with the guidelines laid out in the Research Governance Framework for Health and Social Care (RGF). Failure to do this could lead to permission to carry out this research being withdrawn.
- You must comply with any other relevant guidelines including the Data Protection Act, The Health and Safety Act and local Trust Policies and Guidelines.
- If you encounter any problems during your research you must inform your Sponsor and us immediately to seek appropriate advice or assistance.
- Research projects will be added to any formal Department of Health research register.
- Please complete GCP refresher training

- Please seek advice from Medipex Ltd with regard to any intellectual property generated by the study

Please note that suspected misconduct or fraud should be reported, in the first instance, to local Counter Fraud Specialists for this Trust. R&D staff are also mandated to do this in line with requirements of the RGF.

Adverse incidents relating to the research procedures and/or SUSARs (suspected unexpected serious adverse reactions) should be reported, in line with the protocol requirements, using Trust incident reporting procedures in the first instance and to the chief investigator.

They should also be reported to:

- The R&D Department
- the Research Ethics Committee that gave approval for the study
- other related regulatory bodies as appropriate.

You are required to ensure that all information regarding patients or staff remains secure and *strictly confidential* at all times. You must ensure that you understand and comply with the requirements of the NHS Confidentiality Code of Practice (<http://www.dh.gov.uk/assetRoot/04/06/92/54/04069254.pdf>) and the Data Protection Act 1998. Furthermore you should be aware that under the Act, unauthorised disclosure of information is an offence and such disclosures may lead to prosecution.

Changes to the approved documents **MUST** be approved before any changes can be implemented. Details of changes and copies of revised documents, with appropriate version control, must be provided to the R&D Office. Advice on how to undertake this process can be obtained from R&D.

Projects sponsored by organisations other than the Trust are reminded of those organisations' obligations as defined in the Research Governance Framework, and the requirements to inform all organisations of any non-compliance with that framework or other relevant regulations discovered during the course of the research project.

The research Sponsor or the Chief Investigator, or the local Principal Investigator, may take appropriate urgent safety measures in order to protect research participants against any immediate hazard to their health or safety.

The R&D office should be notified that such measures have been taken. The notification should also include the reasons why the measures were taken and the plan for further action. The R&D office should be notified within the same time frame of notifying the REC and any other regulatory bodies.

Note that NHS indemnities only apply within the limitations of the protocol, and the duties undertaken therewith, by research staff with substantive or honorary research contracts with this Trust.

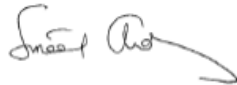
Once you have finished your research you will be required to complete a Project Outcome form. This will be sent to you nearer the end date of your project (Please inform us if the expected end date of your project changes for any reason).

We will require a copy of your final report/peer reviewed papers or any other publications relating to this research. Finally we may also request that you provide us with written

information relating to your work for dissemination to a variety of audiences including service users and carers, members of staff and members of the general public. You must provide this information on request.

If you have any queries during your research please contact us at any time. May I take this opportunity to wish you well with the project.

Yours sincerely

A handwritten signature in black ink, appearing to read 'Sinead Audsley', with a long, sweeping underline that extends to the right.

Sinead Audsley  
Research Governance Manager

CC. Dr Gillian Tober, LYPFT  
Claire Skinner, University of Leeds

### A.3 Regulatory approvals – University ethics committee



UNIVERSITY OF LEEDS

Faculty of Medicine and Health Research Office  
School of Medicine Research Ethics Committee (SoMREC)

Room 9.29, level 9  
Worsley Building  
Clarendon Way  
Leeds, LS2 9NL  
United Kingdom

☎ +44 (0) 113 343 1642

31 October 2016

Ms. Helen Crosby  
PhD Student  
Leeds Institute of Health Sciences  
Faculty of Medicine and Health  
University of Leeds  
Room G.02, Charles Thackrah Building  
101 Clarendon Road  
LEEDS LS2 9LJ

Dear Helen

Ref no: **MREC16-008**

Title: **The Addiction Therapist Rating Scale, a tool for evaluating therapist delivery of psychological therapies used in routine practice for alcohol and drug use problems: A Delphi study**

Your research application has been reviewed by the School of Medicine Ethics Committee (SoMREC) and we can confirm that ethics approval is granted based on the following documentation received from you.

Document	Version	Date Submitted
Helen Crosby_Delphi Study Ethical Review Form	2	28/10/2016
Invitation email to potential participants	2	14/10/2016
Reminder email to potential participants	2	14/10/2016
Bristol Online Surveys (BOS) screenshots: Participant information and consent form	2	14/10/2016

Please notify the committee if you intend to make any amendments to the original research ethics application or documentation. All changes must receive ethics approval prior to implementation. Please contact the Faculty Research Ethics Administrator for further information ([fmhuniethics@leeds.ac.uk](mailto:fmhuniethics@leeds.ac.uk))

Ethics approval does not infer you have the right of access to any member of staff or student or documents and the premises of the University of Leeds. Nor does it imply any right of access to the premises of any other organisation, including clinical areas. The committee takes no responsibility for you gaining access to staff, students and/or premises prior to, during or following your research activities.

Please note: You are expected to keep a record of all your approved documentation, as well as documents such as sample consent forms, and other documents relating to the study. This should be kept in your study file, which should be readily available for audit purposes. You will be given a two week notice period if your project is to be audited.

It is our policy to remind everyone that it is your responsibility to comply with Health and Safety, Data Protection and any other legal and/or professional guidelines there may be.

We wish you every success with the project.

Yours sincerely

A handwritten signature in black ink, appearing to read 'Naomi Quinton', with a horizontal line underneath.

**Co-Chair, SoMREC, University of Leeds**

*(Approval granted by Co-Chair Dr Naomi Quinton on behalf of committee)*

## A.4 Regulatory approvals – Amendment 1

### NRES Committee Yorkshire & The Humber - Leeds West

Room 001, Jarrow Business Centre  
Rolling Mill Road  
Jarrow  
Tyne and Wear  
NE32 3DT

Tel: 0191 428 3444  
Fax:

22 May 2015

Miss Helen F. Crosby  
Researcher/PhD Student  
University of Leeds  
Leeds Institute of Health Sciences  
Room G.02, Charles Thackrah Building  
101 Clarendon Road  
Leeds  
LS2 9NG

Dear Miss Crosby

**Study title:** The development of the Addiction Therapist Rating Scale (ATRS): a rating scale for evaluating the delivery of evidence based treatments used in routine practice for alcohol and illicit drug use problems.

**REC reference:** 15/YH/0037

**Amendment number:** Substantial Amendment 1

**Amendment date:** 23 April 2015

**IRAS project ID:** 165130

The above amendment was reviewed by the Sub-Committee in correspondence.

#### Summary of amendment

This amendment was submitted to include service users' friends or family members on the audio and visual recordings from routine practice data.

#### Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

#### Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Notice of Substantial Amendment (non-CTIMP)	Substantial Amendment 1	23 April 2015
Participant consent form [Study 2a Treatment Session Recordings FFM Consent Form]	1	23 April 2015

A Research Ethics Committee established by the Health Research Authority

Participant information sheet (PIS) [Study 2a Treatment Session Recordings FFM Information Sheet]	1	23 April 2015
Research protocol or project proposal	3	10 April 2015

#### Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

#### R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

#### Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at <http://www.hra.nhs.uk/hra-training/>

15/YH/0037:	Please quote this number on all correspondence
-------------	--

Yours sincerely  
pp



Dr Sheila E. Fisher  
Chair

E-mail: [nrescommittee.yorkandhumber-leedswest@nhs.net](mailto:nrescommittee.yorkandhumber-leedswest@nhs.net)

*Enclosures:* List of names and professions of members who took part in the review

*Copy to:* Ms Sinead Audsley, Leeds and York Partnership NHS Foundation Trust

Clare Skinner, University of Leeds

## A.5 Regulatory approvals – Amendment 2



### Health Research Authority

NRES Committee Yorkshire & The Humber - Leeds West

Room 001, Jarrow Business Centre  
Rolling Mill Road  
Jarrow  
Tyne and Wear  
NE32 3DT

Tel: 0191 4283548

14 September 2015

Miss Helen F. Crosby  
Researcher/PhD Student  
University of Leeds  
Leeds Institute of Health Sciences  
Room G.02, Charles Thackrah Building  
101 Clarendon Road  
Leeds  
LS2 9NG

Dear Miss Crosby

**Study title:** The development of the Addiction Therapist Rating Scale (ATRS): a rating scale for evaluating the delivery of evidence based treatments used in routine practice for alcohol and illicit drug use problems.

**REC reference:** 15/YH/0037

**Amendment number:** Substantial Amendment 2

**Amendment date:** 17 August 2105

**IRAS project ID:** 165130

The above amendment was reviewed by the Sub-Committee in correspondence.

#### Summary of Amendment

This amendment was submitted to inform the Committee of the addition of a new site to the study to aid with recruitment.

Furthermore, the aim within the study was to collect 50 recordings from one recruitment site by June 2015. However, the specialist NHS addictions service underwent a recommissioning and restructuring process. As a result, therapists were not able to prioritise the research and only 18 useable recordings were collected.

#### Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

#### Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
----------	---------	------



Notice of Substantial Amendment (non-CTIMP)	Substantial Amendment 2	17 August 2105
Participant information sheet (PIS) [Study 2a Treatment Session Recordings Wales Service Uder Information Sheet]	1	26 August 2015
Participant information sheet (PIS) [Study 2a Treatment Session Recordings Wales Therapist Information Sheet]	1	26 August 2015
Research protocol or project proposal [The ATRS Protocol]	4	21 July 2015

#### Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

#### R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

#### Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at <http://www.hra.nhs.uk/hra-training/>

<b>15/YH/0037:</b>	<b>Please quote this number on all correspondence</b>
--------------------	---

Yours sincerely



pp

**Mr Anthony Warnock-Smith**  
Alternate Vice Chair

E-mail: [nrescommittee.yorkandhumber-leedswest@nhs.net](mailto:nrescommittee.yorkandhumber-leedswest@nhs.net)

*Enclosures: List of names and professions of members who took part in the review*

*Copy to: Ms Sinead Audsley, Leeds and York Partnership NHS Foundation Trust*

*Ms Clare Skinner, University of Leeds*

## **Appendix B**

### **Chapter 3 Appendices**

Appendix B focuses on the appendices for chapter 3, identifying existing measures. The appendices are as follows:

- i) Search strategy for identifying existing fidelity measures (p.209).
- ii) Summary of fidelity measures and associated development and/or validation articles excluded from the review (p.212).
- iii) Summary of therapies covered by the identified measures (p.220).
- iv) Summary of the development and validation of the identified measures (p.224).
- v) Methods of establishing construct validity (p.238).
- vi) Variants of the intraclass correlation coefficient (p.239).

## B.1 Search strategy for identifying existing fidelity measures

The search strategy for each of the three databases is listed below:

### Embase (Embase Classic+Embase, 1947 to January 2015)

1. (treatment\* adj3 fidelity).tw
2. (treatment\* adj3 integrity).tw
3. "fidelity measur\*".tw
4. (intervention\* adj3 fidelity).tw
5. (fidelity adj3 monitor\*).tw
6. "process rating".tw
7. (therapist adj3 adherence).tw
8. "practitioner adherence".tw
9. "clinician adherence".tw
10. (therapist\* adj3 competence).tw
11. (practitioner\* adj3 competence).tw
12. (clinician\* adj3 competence).tw
13. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
14. "rating scale\*".ti
15. "scale\*".ti
16. "instrument\*".ti
17. "measur\*".ti
18. "checklist\*".ti
19. "assess\*".ti
20. rating scale/
21. psychological rating scale/
22. measurement/
23. checklist/
24. 14 or 15 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23
25. "validity".ti
26. "reliability".ti
27. "psychometric\*".ti
28. "validation".ti
29. exp validity/
30. exp reliability/
31. validation study/
32. instrument validation/
33. 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32
34. 24 or 33
35. 13 and 34
36. "child\*.tw"
37. "adoles\*.tw"
38. 36 or 37
39. 35 not 38

## Medline (1946 to January 2015)

1. (treatment\* adj3 fidelity).tw
2. (treatment\* adj3 integrity).tw
3. "fidelity measur\*".tw
4. (intervention\* adj3 fidelity).tw
5. (fidelity adj3 monitor\*).tw
6. "process rating".tw
7. (therapist adj3 adherence).tw
8. "practitioner adherence".tw
9. "clinician adherence".tw
10. (therapist\* adj3 competence).tw
11. (practitioner\* adj3 competence).tw
12. (clinician\* adj3 competence).tw
13. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
  
14. "rating scale\*".ti
15. "scale\*".ti
16. "instrument\*".ti
17. "measur\*".ti
18. "checklist\*".ti
19. "assess\*".ti
20. 14 or 15 or 15 or 16 or 17 or 18 or 19
  
21. "validity".ti
22. "reliability".ti
23. "psychometric\*".ti
24. "validation".ti
25. Validation studies/
26. Psychometrics/
27. 21 or 22 or 23 or 24 or 25 or 26
  
28. 20 or 27
29. 13 and 28
  
30. "child\*.tw"
31. "adoles\*.tw"
32. 30 or 31
  
33. 29 not 32

**PsycINFO (1806 to January 2015)**

1. (treatment\* adj3 fidelity).tw
2. (treatment\* adj3 integrity).tw
3. "fidelity measur\*".tw
4. (intervention\* adj3 fidelity).tw
5. (fidelity adj3 monitor\*).tw
6. "process rating".tw
7. (therapist adj3 adherence).tw
8. "practitioner adherence".tw
9. "clinician adherence".tw
10. (therapist\* adj3 competence).tw
11. (practitioner\* adj3 competence).tw
12. (clinician\* adj3 competence).tw
13. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
  
14. "rating scale\*".ti
15. "scale\*".ti
16. "instrument\*".ti
17. "measur\*".ti
18. "checklist\*".ti
19. "assess\*".ti
20. exp Rating Scales/
21. Measurement/
22. 14 or 15 or 15 or 16 or 17 or 18 or 19 or 19 or 21
  
23. "validity".ti
24. "reliability".ti
25. "psychometric\*".ti
26. "validation".ti
27. exp Test Construction/
28. Psychometrics/
29. 23 or 24 or 25 or 26 or 27 or 28
  
30. 22 or 29
31. 13 and 29
  
32. "child\*.tw"
33. "adoles\*.tw"
34. 32 or 33
  
35. 31 not 34

## B.2 Summary of fidelity measures and associated development and/or validation articles excluded from the literature review

Measure (Author, Year)	Clinical Area	Target Therapy	Assessment of Fidelity	No. Items	Item scoring	Article Author (Year)	Validation	Copy*	Reason for Exclusion
1. ACF (Brauhardt and Hilbert, 2010)	Binge-eating disorders	Cognitive behavioural therapy (CBT)	Adherence	10	3-point Likert-type scales scoring item adherence	Brauhardt et al. (2014)	Reliability & validity analyses	No	Unable to obtain measure
2. BTST-1 (Amini and Woolley, 2011)	Behaviour problems and drug use	Brief strategic family therapy (BSFT)	Proficiency	Not clear	5-point Likert-type scales scoring proficiency of behaviours	Amini and Woolley (2011)	No	Yes	Family therapy
3. CTCS-SP (Clark et al., 2007)	Social phobia	Cognitive therapy (CT)	Competence	16	7-point Likert-type scales scoring competence	von Consbruch et al. (2012)	Reliability analyses	No	Unable to obtain measure
4. CTPAS (Startup et al., 2002)	Psychosis	CBT	Adherence (extensiveness)	12	7-point Likert-type scales scoring item extensiveness	Startup et al. (2002)	Reliability & validity analyses	No	Treatment of psychosis too removed from the addiction field
5. FIPAS (Onwumere et al., 2009)	Psychosis	Family intervention in psychosis	Adherence (frequency)	14	8-point Likert-type scales scoring item frequency	Onwumere et al. (2009)	Reliability analysis	No	Family therapy
6. FRASE (Hatch-Maillette et al., 2013)	HIV and substance use	Real Men Are Safe – Culturally Adapted (REMAS-CA)	Adherence (extensiveness), competence (therapist skill)	22	5-point Likert-type scales scoring item extensiveness and therapist skill	Hatch-Maillette et al. (2013)	Reliability analyses	No	Group therapy

Measure (Author, Year)	Clinical Area	Target Therapy	Assessment of Fidelity	No. Items	Item Scoring	Article Author (Year)	Validation	Copy	Reason for Exclusion
7. ICS (Tadic and Despland, 2001)	Anxiety, mood or personality disorders	Brief psychodynamic investigation (BPI)	Competence (therapist skill)	33	5-point Likert scales scoring therapist skill	Tadic et al. (2003)	Reliability & validity analyses	No	BPI not typically used in the addiction field and not adapted for alcohol or drug use problems
8. IT-IS (McGuire et al., 2012)	Severe mental illness	Illness management and recovery (IMR)	Competence	16	5-point Likert scales scoring competence	McGuire et al. (2012)	Reliability & validity analyses	Yes	Treatment of schizophrenia too removed from the addiction field
9. MAC (Collins et al., 2009)	Smoking cessation	Smoking cessation interventions	Adherence	99	Dichotomous scoring ("present/not present")	Collins et al. (2009)	Reliability & validity analyses	No	Not written in English (German)
10. MBCT-AS (Segal et al., 2002)	Depression	Mindfulness- based cognitive therapy (MBCT)	Adherence (evidence for items)	17	3-point Likert scales scoring level of evidence for items	Segal et al. (2002)	Reliability & validity analyses	No	Group therapy
11. MBRP-AC (Chawla et al., 2010)	Substance use	Mindfulness- based relapse prevention (MBRP)	Adherence (checklist and frequency), competence	19	Dichotomous scoring (7-items), behaviour counts (4-items), 5-point Likert scales scoring competence (8-items)	Chawla et al. (2010)	Reliability analyses	Yes	Group therapy

Measure (Author, Year)	Clinical Area	Target Therapy	Assessment of Fidelity	No. Items	Item Scoring	Article Author (Year)	Validation	Copy	Reason for Exclusion
12. MBT-ACS (Karterud and Bateman, 2010)	Borderline personality disorder (BPD)	Mentalization based treatment (MBT)	Adherence (frequency and extensiveness), competence (therapist skill)	17	7-point Likert-type scales scoring item frequency, extensiveness, and therapist skill	Karterud et al. (2013)	Reliability & validity analyses	No	Treatment of BPD too removed from the addiction field
13. MCP-AS (Prowse and Nagel, 2013)	Depression, substance use	Motivational care planning (MCP)	Adherence (evidence for items)	10	9-point Likert scales scoring level of evidence for items	Prowse and Nagel (2014)	No	No	Unable to obtain measure
14. MISC (Miller et al., 2008a)	Behaviour change	Motivational interviewing (MI)	Competence (quality)	N/A	Global ratings of therapist behaviours, coding and counting therapist-client utterances	Moyers et al. (2003)	Reliability analyses	Yes	Not a fidelity measure (a coding system that quantifies therapist-client interactions)
15. MITI (Moyers et al., 2015)	Behaviour change	MI	Therapist competence	N/A	5-point Likert-type scales scoring four global dimensions, coding therapist utterances (behaviour counts)	Moyers et al. (2005)	Reliability & validity analyses	Yes	Not a fidelity measure (a behavioural coding system)
16. MatchTRS (Carroll et al., 1998a)	Substance use	CBT, motivational enhancement therapy (MET), twelve-step facilitation (TSF)	Adherence	Not clear	Likert-type scales	Carroll et al. (1998a)	Reliability & validity analyses	No	Unable to obtain (measure no longer exists, incorporated into YACS)



Measure (Author, Year)	Clinical Area	Target Therapy	Assessment of Fidelity	No. Items	Item Scoring	Article Author (Year)	Validation	Copy	Reason for Exclusion
17. PACS-SE (Barber, 1988)	Depression	Supportive-expressive psychotherapy (SEC)	Adherence (frequency), competence (quality)	45	7-point Likert-type scales scoring item frequency and quality	Barber and Crits-Christoph (1996)	Reliability & validity analyses	No	SEC not typically used in the addiction field and not adapted for alcohol or drug use problems
18. PPPCQ (Stein et al., 2007)	Eating disorders	Identity intervention program (IIP)	Treatment fidelity	98	Dichotomous scoring ("present/absent")	Stein et al. (2007)	Reliability & validity analyses	No	IIP not typically used in the addiction field and not adapted for alcohol or drug use problems
19. PPRS-BPD (Levy et al., 2006)	BPD	Dialectical behavioural therapy (DBT), transference-focused therapy (TFP), supportive psychotherapy (SPT)	Adherence (frequency), competence (quality)	238	9-point Likert-type scales scoring item frequency and quality	Levy et al. (2006)	Reliability analysis	No	Treatment of BPD too removed from the addiction field
20. PQS (Jones, 2000)	Mental health	Psychodynamic and cognitive behavioural therapies	Fidelity not assessed	100	N/A	Jones and Pulos (1993)	No	Yes	Not a fidelity measure (focuses on the therapist-client interaction)

Measure (Author, Year)	Clinical Area	Target Therapy	Assessment of Fidelity	No. Items	Item scoring	Article Author (Year)	Validation	Copy	Reason for Exclusion
21. PST-PAC (Hegel et al., 2004)	Depression	Problem-solving treatment	Competence (therapist skill)	7	6-point Likert scales scoring therapist skill	Hegel et al. (2004)	Reliability analyses	No	Unable to obtain measure
22. ROSTA (Hartley et al., 2014)	Psychosis	CBT	Appropriateness	Not clear	3-point Likert-type scale scoring item appropriateness	Hartley et al. (2014)	Reliability & validity analyses	No	Therapy delivered by telephone
23. RPT-FS (Alvarez- Jimenez et al., 2008)	Psychosis	Relapse prevention therapy (RPT)	Rater identifies phase of therapy (%), minimum cut-off scores for therapy phases and general therapeutic factors	45	Unclear	Alvarez- Jimenez et al. (2008)	Reliability & validity analyses	No	Treatment of psychosis too removed from the addiction field
24. SPIRIT Treatment Fidelity Assessment Tool (Song et al., 2010)	End stage renal disease	Sharing the Patient's Illness Representations to Increase Trust (SPIRIT) intervention	Adherence, and competence	Not clear	Scoring elements of intervention on a 4- point scale, counting occurrences of process skills, assessing pacing, 3-point Likert scales scoring participant responsiveness	Song et al. (2010)	Reliability analyses	No	Treatment for end stage renal disease too removed from the addiction field
25. SEA-FM (Allan and Ungar, 2014)	Families with complex needs	Social Ecological Approach (SEA)	Fidelity (therapist skill)	17	5-point Likert scales scoring therapist skill	Allan and Ungar (2014)	No	No	Family therapy

Measure (Author, Year)	Clinical Area	Target Therapy	Assessment of Fidelity	No. Items	Item Scoring	Article Author (Year)	Validation	Copy	Reason for Exclusion
26. STCF (Svartberg, 1989)	Psychotherapy techniques	Short-term anxiety-provoking psychotherapy (STAPP)	Competence	11	5-point Likert scales scoring STAPP competence	Svartberg (1989)	Reliability & validity analyses	No	Unable to obtain measure
27. TAS (Hoyt et al., 1981)	Psychotherapy techniques	Dynamic, short-term psychotherapy	Assessing therapist actions	25	5-point Likert scales scoring therapist actions	Hoyt et al. (1981)	Reliability & validity analyses	Yes	Dynamic therapy not typically used in the addiction field and not adapted for alcohol or drug use problems
28. TAC (Barton et al., 2008)	Depression	CBT	Adherence (extensiveness)	30	5-point Likert-type scales scoring item extensiveness	Barton et al. (2008)	Reliability analysis	No	Unable to obtain measure
29. TSRF (O'Malley et al., 1988)	Depression	Interpersonal psychotherapy (IPT)	Competence (therapist skill)	Not clear	7-point Likert scales scoring therapist skill	O'Malley et al. (1988)	No	No	IPT not typically used in the addiction field and not adapted for alcohol or drug use problems
30. Unnamed Scale-1 (Huppert et al., 2001)	Panic disorder	CBT	Adherence (extensiveness), competence (therapist skill)	Not clear	7-point Likert-type scales scoring extensiveness and global ratings of therapist skill	Huppert et al. (2001)	No	No	Unable to obtain measure

Measure (Author, Year)	Clinical Area	Target Therapy	Assessment of Fidelity	No. Items	Item Scoring	Article Author (Year)	Validation	Copy	Reason for Exclusion
31. Unnamed Scale-2 (Boswell et al., 2013)	Panic disorder	CBT	Adherence, competence	Not clear	Scoring of adherence unclear, 5-point Likert scale scoring how well session goal was accomplished.	Boswell et al. (2013)	Reliability analysis	No	Unable to obtain measure
32. Unnamed Scale-3 (Strang and McCambridge, 2004)	Substance use	MI	Characteristics of MI, including quantity of change talk	Not clear	Items scored dichotomously, and categorical options on 4-point, 5-point, and 9-point Likert scales	Strang and McCambridge (2004)	No	No	Unable to obtain measure
33. Unnamed Scale-4 (Godfrey et al., 2007)	Chronic fatigue in primary care	CBT and counselling	Adherence (extensiveness)	14	7-point Likert-type scales scoring item extensiveness	Godfrey et al. (2007)	Reliability & validity analyses	No	Unable to obtain measure
34. VASE-R (Rosengren et al., 2009)	Substance use	MI	Therapist skill	18	N/A	Rosengren et al. (2008)	Reliability & validity analyses	Yes	Not a fidelity measure (a coding framework using video-based vignettes)

\*Copy = Copy of the measure obtained; No. Items = Number of items; N/A = Not applicable.

**ACF** = Adherence Control Form; **BTST-1** = Brief Strategic Therapy Scale-1; **CTCS-SP** = Cognitive Therapy Competence Scale for Social Phobia; **CTPAS** = Cognitive Therapy for Psychosis Adherence Scale; **FIPAS** = Family Intervention in Psychosis – Adherence Scale; **FRASE** = Fidelity Rating and Skill Evaluation; **ICS** = Investigation Competence Scale; **IT-IS** = Illness Management and Recovery Treatment Integrity Scale; **MAC** = Manual Adherence Checklist; **MBCT-AS** = Mindfulness-Based Cognitive Therapy Adherence Scale; **MBRP-AC** = Mindfulness-Based Relapse Prevention Adherence and Competence Scale; **MBT-ACS** = Mentalization-Based Treatment (MBT) Adherence and Competence Scale; **MCP-AS** = Motivational Care Planning-Adherence Scale; **MISC** = Motivational Interviewing Skills Code Version 2.1; **MITI** = Motivational Interviewing Treatment Integrity; **MatchTRS** = MATCH Tape Rater Scale; **PACS-SE** = Penn Adherence/ Competence Scale for Supportive-Expressive (SE) Dynamic Psychotherapy; **PPPCQ**

= Possibilities Project Psychotherapy Coding Questionnaire; **PPRS-BPD** = Psychotherapy Process Rating Scale for Borderline Personality Disorder; **PQS** = Psychotherapy Process Q-Set; **PST-PAC** = Problem-solving Treatment Adherence and Competence Scale; **ROSTA** = Recovery Orientated CBT for psychosis: supported Self-help and Telephone therapy Adherence scale; **RPT-FS** = Relapse Prevention Therapy-Fidelity Scale; **SPIRIT Treatment Fidelity Assessment Tool** = Sharing the Patient's Illness Representations to Increase Trust (SPIRIT) Treatment Fidelity Assessment Tool; **SEA-FM** = Social Ecological Approach – Fidelity Measure; **STCRF** = Short-Term Anxiety Provoking Psychotherapy (STAPP) Therapist Competence Rating Form; **TAS** = Therapist Action Scale; **TAC** = Treatment Adherence Checklist; **TSRF** = Therapist Strategy Rating Form; **Unnamed Scale-1** = Unnamed Adherence and Competence Scale No.1 Rating Cognitive Behavioural Therapy; **Unnamed Scale-2** = Unnamed Adherence and Competence Scale No.2 Rating Cognitive Behavioural Therapy; **Unnamed Scale-3** = Unnamed Scale No.3 Rating Motivational Interviewing; **Unnamed Scale-4** = Unnamed Scale No.4 to Assess Treatment Fidelity and Predict Outcome; **VASE-R** = Video Assessment of Simulated Encounters Revised

### B.3 Summary of therapies covered by the identified fidelity measures

Therapy	Therapy Description	Associated Measure(s)
An alcohol focused intervention (AF)	A manual-guided adaptation of social behaviour and network therapy (SBNT; Copello et al., 2002), which helps clients to build networks of people supportive of positive change in drinking (UKATT Research Team, 2001; Watson et al., 2015).	ADAPTA PRS
Brief advice (BA)	A five-minute session of brief advice, involving feedback of the results of the screening, and discussion on the health consequences of continued hazardous alcohol use (Watson et al., 2013a).	AESOPS PRS
Behaviour change counselling (BCC)	A brief adaptation of motivational interviewing (MI; Miller and Rollnick, 2013) for use in healthcare settings. BCC involves addressing clients' motivation to change health-related behaviours and discussing a plan for change (Watson et al., 2013a; Lane et al., 2005).	AESOPS PRS, and BECCI
Brief negotiation interview (BNI)	A ten-minute adaptation of MI (Miller and Rollnick, 2013) for use in healthcare settings, combined with physician advice and behavioural contracting (D'Onofrio et al., 2005; Pantaloni et al., 2012).	BAS
Cognitive-behavioural (CB) treatments	A group of therapies broadly defined to include: cognitive, cognitive-behavioural, and behavioural therapies (Hilsenroth et al., 2005).	CPPS
Cognitive-behavioural therapy (CBT)	<p>CBT, in this review, refers to three manual-guided, time-limited applications of CBT for specific problems:</p> <p>i) A form of CBT for problematic substance use based on the principles of social learning theory; substance use behaviours were viewed "<i>as functionally related to major problems in the individual's life</i>" (Carroll et al., 1998a, p.292). The approach focuses on exploring clients' substance use, including identifying high risk situations, and coping skills training for problems, such as, social pressure to drink, and depression (Carroll, 1997; Carroll et al., 1998a; Carroll et al., 2000; Karno and Longabaugh, 2007).</p> <p>ii) A form of CBT for depression based on Beck's cognitive therapy model (Moorey, 2002). The underlying principle is that negative emotions and maladaptive behaviours are influenced by distorted cognitions and patterns of thinking (Gaudiano, 2008). The approach aims to identify, evaluate, reality-test, and modify dysfunctional beliefs (schemas) underlying the distorted cognitions (Elkin et al., 1985). Behavioural techniques were also included as part of the therapy (DeRubeis et al., 1982; Elkin et al., 1985; Hill et al., 1992).</p> <p>iii) A form of CBT for post-traumatic stress disorder (PTSD), which includes: "<i>breathing retraining</i> [anxiety management skills training] <i>and education about PTSD, with a primary focus on teaching cognitive restructuring to address core beliefs related to trauma that are thought to underlie PTSD symptoms</i>" (Lu et al., 2012, p.786).</p>	<p>CBT Therapist Checklist, TPRS, and YACSII</p> <p>CSPRS-6, and MTRS</p> <p>CBT for PTSD Fidelity Scale</p>

Therapy	Therapy Description	Associated Measure(s)
Clinical Management (ClinM)	A low-intensity treatment for drug dependence, adapted from guidelines developed for the NIMH Treatment of Depression Collaborative Research Program (Fawcett et al., 1987; Elkin et al., 1985). ClinM, also called compliance enhancement, provides medication management without using active ingredients specific to other therapies for drug use problems (Carroll et al., 1999; Carroll et al., 2000).	CSPRS-6, and YACSII
Cognitive therapy (CT)	A collaborative, goal-orientated therapy based on Beck's cognitive theory (Lynn, 2015). Similar to the form of CBT for depression described above (see CBT), the underlying principle of CT is that negative emotions and maladaptive behaviours are influenced by distorted cognitions and patterns of thinking (Gaudiano, 2008). The approach aims to identify, evaluate, reality-test, and modify dysfunctional beliefs (schemas) underlying the distorted cognitions (Elkin et al., 1985). Behavioural techniques can be included as part of therapy.	CTACS, and CTS-R
Compliance enhancement (CE)	See ClinM.	CE Therapist Rating Form
Contingency management (CM)	A behaviour modification intervention that uses motivational incentives to reinforce client behaviours. The treatment manual associated with the CM Clinician Rating form describes the use of contingency management to target abstinence from drugs and treatment attendance (Petry and Stitzer, 2002).	CM Clinician Rating Form
Exploratory therapy (ET)	A manual-guided therapy for depression and/or anxiety based on psychodynamic and experiential principles. The relationship-orientated approach enables clients to experience feelings and conflicts in the therapy session, and uses the therapeutic relationship to explore and understand the clients' experiences (Barkham et al., 1989; Startup and Shapiro, 1993).	SPRS
A health living intervention (HL)	A manual-guided intervention based on the principles of behaviour change counselling (BCC). Clients chose to change their behaviour in up to three health behaviour domains from a choice of seven, including: drinking, drug use, diet, smoking, exercise, personal care, and medication compliance (Watson et al., 2015).	ADAPTA PRS
I Am Concerned (IAC) brief opportunistic intervention	A structured brief intervention for use in prenatal interviews (meetings clients have with their healthcare provider during pregnancy) following client disclosure of substance use. The abstinence-orientated intervention provides educational materials, including photographs of children exposed to in utero substance use, and encourages referral to addiction services (Torrey, 2011; Torrey, 2012).	IAC Treatment Fidelity Instrument
Individual drug counselling (IDC)	A manual-guided abstinence-orientated form of drug counseling for cocaine dependence. Consistent with the twelve-step approach, IDC focuses on establishing motivation, early abstinence, and maintaining abstinence (Barber et al., 1996).	ACS-IDCCD

Therapy	Therapy Description	Associated Measure(s)
Integrated motivational interviewing and cognitive-behavioural therapy (MI-CBT)	A form of CBT for chronic treatment-resistant psychosis combined with a modified version of MI (Miller and Rollnick, 2013) for substance use problems (Lu et al., 2012). The integrated approach focuses on building and reinforcing motivation to change, and developing a plan for change (Barrowclough et al., 2010).	MI-CTS
Interpersonal therapy (IPT)	A structured, time-limited therapy that is based on the principle that psychiatric disorders, including depression and substance use, are associated with interpersonal difficulties. The approach aims to help clients: i) achieve a reduction in depressive symptoms/reduce or abstain from substance use, and ii) develop strategies for dealing with the interpersonal difficulties associated with clients' depression, or substance use (Weissman and Klerman, 2015; Nuro et al., 2005).	CSPRS-6, MTRS, and YACSII
Manual assisted cognitive therapy (MACT)	A brief form of CT for use with clients immediately after an episode of self-harm (parasuicide). The main components include: an evaluation of the self-harm attempt, crisis skills, problem solving, management of emotions and negative thinking, and relapse prevention (Tyrer et al., 2003; Davidson et al., 2004).	MACT Rating Scale
Motivational interviewing (MI)	A person-centred, goal-orientated therapy for eliciting and strengthening motivation to change substance use problems and other problem areas (Wampold and Imel, 2015). MI uses a collaborative and empathic approach, also referred to as a MI style, to help clients explore and resolve ambivalence about change (Norcross, 2010; Miller and Rollnick, 2013).	GROMIT, ITRS, MISTS Revised <sup>1</sup> , and YACSII
Motivational Enhancement Therapy (MET)	A manual-based adaptation of MI (Miller and Rollnick, 2013), which reinforces clients' motivation for and commitment to change their drinking (Miller et al., 1992; Watson et al., 2013a; Tober et al., 2008).	AESOPS PRS, and UKATT PRS
Prescriptive therapy (PT)	A form of CBT for depression that involves: anxiety management (e.g., relaxation, and graded exposure), self-management procedures (e.g., self-monitoring, and assertion training), cognitive restructuring (e.g. challenging dysfunctional thoughts), and job-strain management (e.g., time management, and remediation of memory problems) (Barkham et al., 1989; Shapiro and Startup, 1992).	SPRS
Psychodynamic-interpersonal (PI) treatments	A group of therapies broadly defined to include: psychodynamic, psychodynamic-interpersonal, and interpersonal therapies (Hilsenroth et al., 2005).	CPPS
Psychotherapy	A generic term used to describe " <i>primarily interpersonal treatment that is a) based on psychological principles; b) involves a trained therapist and a client who is seeking help for a mental disorder, problem, or complaint; c) is intended by the therapist to be remedial for the client disorder, problem, or complaint; and d) is adapted or individualized for the particular client and his or her disorder, problem, or complaint.</i> " (Wampold and Imel, 2015, p.37)	TPRS



Therapy	Therapy Description	Associated Measure(s)
Social behaviour and network therapy (SBNT)	A manual-based therapy, using cognitive and behavioural strategies, to help clients build networks of people supportive of positive change in drinking and associated behaviours (UKATT Research Team, 2001; Copello et al., 2002; Tober et al., 2008).	UKATT PRS
Supportive-expressive therapy (SET)	A manualised short-term psychodynamic treatment for cocaine use. SET integrates biological and psychological perspectives to help clients gain an understanding of “ <i>conflictual relationship patterns in the context of a supportive relationship</i> ” (Connolly et al., 1998, p.291; Crits-Christoph et al., 2008).	ACS-SEC
Twelve-step facilitation (TSF)	A manual-guided approach that seeks to encourage client abstinence from psychoactive drugs and increase clients’ participation in self-help groups (Alcoholics Anonymous, Cocaine Anonymous, Narcotics Anonymous) outside therapy sessions (Carroll et al., 2000; Campbell et al., 2013).	TSF-ACES, and YACSII

<sup>1</sup> MISTS Revised evaluates the delivery of motivational interviewing and its derivatives.

NIMH = National Institute of Mental Health.

**ACS-IDCCD** = Adherence/Competence Scale for Individual Drug Counseling for Cocaine Dependence; **ACS-SEC** = Adherence/Competence Scale for Supportive-Expressive Therapy for Cocaine Dependence; **ADAPTA PRS** = ADAPTA (Addressing Drinking Among Patients: comparing Two Approaches) Process Rating Scale; **AESOPS PRS** = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale; **BAS** = Brief Negotiation Interview Adherence Scale; **BECCI** = Behaviour Change Counselling Index; **CBT for PTSD Fidelity Scale** = Cognitive-Behavioral Therapy for Post-Traumatic Stress Disorder Fidelity Scale; **CBT Therapist Checklist** = Cognitive Behavioral Therapy Therapist Checklist; **CE Therapist Rating Form** = Compliance Enhancement Therapist Adherence/Competence Rating Form; **CM Clinician Rating Form** = Contingency Management Clinician Adherence/Competence Rating Form; **CPPS** = Comparative Psychotherapy Process Scale; **CSPRS-6** = Collaborative Study Psychotherapy Rating Scale – Form 6; **CTACS** = Cognitive Therapy Adherence and Competence Scale; **CTS-R** = Cognitive Therapy Scale – Revised; **GROMIT** = Global Rating of Motivational Interviewing Therapist; **IAC Treatment Fidelity Instrument** = I Am Concerned Treatment Fidelity Instrument; **ITRS** = Independent Tape Rater Scale; **MACT Rating Scale** = Manual Assisted Cognitive Therapy Rating Scale; **MI-CTS** = Integrated Motivational Interviewing and Cognitive Behavioural Therapy Fidelity Scale; **MISTS Revised** = Motivational Interviewing Supervision and Training Scale Revised; **MTRS** = Minnesota Therapy Rating Scale; **SPRS** = Sheffield Psychotherapy Rating Scale; **TPRS** = Therapy Process Rating Scale; **TSF-ACES** = Twelve Step Facilitation Adherence Competence Empathy Scale; **UKATT PRS** = UK Alcohol Treatment Trial Process Rating Scale; **YACSII** = Yale Adherence and Competence Scale Second Edition.

## B.4 Summary of the development and validation of the identified measures

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of Measure	Validation Data	Main Validation Analyses
1. ACS-IDCCD (Barber et al., 1996)	Primary diagnosis of cocaine dependence	Individual drug counselling (IDC), supportive-expressive therapy (SET), and cognitive therapy (CT)	Derived from the IDC treatment manual	62 audio recordings (41 IDC, 11 CT, and 10 SET)	<p><i>Reliability:</i> Intraclass correlation coefficients (ICCs) (2,3) for the total scale score and five subscales, separated for adherence and competence, ranged from 0.55 to 0.89, indicating fair to excellent reliability among the three raters' scores. Cronbach's alpha (<math>\alpha</math>) for the total and subscale scores ranged from 0.43 to 0.95; acceptable internal consistency (&gt;0.7) for all but one subscale.</p> <p><i>Discriminant validity:</i> Adherence and competence ratings for the total scale score and subscales were compared between therapies using contrast analyses, respective effect sizes (Cohen's <math>d^1</math>) were presented. IDC therapists generally used IDC techniques more frequently and competently than SE and CT therapists (IDC vs. SE: <math>d = 1.01</math> to <math>2.08</math>; IDC vs. CT: <math>d = 0.003</math> to <math>1.07</math>).</p>
2. ACS-SEC (Barber et al., 1997)	Primary diagnosis of cocaine dependence	SET, CT, and IDC	Adapted from the Penn Adherence-Competence Scale for Supportive-Expressive therapy (PACS-SE; Barber and Crits-Christoph, 1996). Items revised using SET and cocaine addiction treatment manuals.	52 audio recordings (32 SET, 10 CT, and 10 IDC)	<p><i>Reliability:</i> ICCs (2,2) for adherence, appropriateness, and quality for the total scale and three subscales ranged from 0.28 to 0.89; reliability was acceptable for adherence (ICCs: 0.61 to 0.89) but poor for appropriateness (ICCs: 0.29 to 0.60) and quality (ICCs: 0.28 to 0.55). Cronbach's alpha for the total scale and subscales ranged from 0.72 to 0.95 demonstrating acceptable internal consistency.</p> <p><i>Discriminant validity:</i> t-tests were used to compare the three therapies. SE therapists used more expressive techniques than CT or IDC therapists, and more supportive techniques than IDC therapists, these differences were statistically significant. Statistical differences were not found between SE and the other types of therapy for the cocaine subscale.</p>

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
3. ADAPTA PRS (Watson et al., 2013b)	-	-	Adapted from the UKATT PRS <sup>(No.25)</sup> and ITRS <sup>(No.17)</sup> .	-	N/A: Development of the measure referred to in the ADAPTA trial protocol (Watson et al., 2013b).
4. AESOPS PRS (Watson et al., 2013a)	Older hazardous alcohol users in primary care	Brief advice (BA), and behaviour change counselling (BCC)	Adapted from the UKATT PRS <sup>(No.25)</sup> and ITRS <sup>(No.17)</sup> .	160 audio recordings (79 BA, and 81 BCC)	<p><i>Reliability:</i> ICCs (3,2) for task and style summary measures, separated for frequency and quality, ranged from 0.64 to 0.81, showing good to excellent reliability among the two raters' scores. The confidence intervals (CIs) were relatively wide, particularly for the quality ratings, indicating uncertainty about the magnitude of the effect.</p> <p><i>Discriminant validity:</i> A linear regression model showed statistically significant differences between BA and BCC; the BCC sessions had significantly higher task and style scores, for both frequency and quality, than the BA session, indicating that the two therapies were distinct.</p>
5. BAS (Pantalon et al., 2012)	Hazardous and harmful drinking among emergency department patients	Brief Negotiation Interview (BNI), and discharge instructions (DI)	Items generated using the BNI treatment manual. Items reduced and revised following item piloting.	342 audio recordings (165 BNI, and 177 DI)	<p><i>Reliability:</i> ICCs (3,3) for 18 of the BAS items ranged from 0.03 to 1.00; 15 items showed good to excellent reliability (ICCs: 0.67 to 1.0), and three items showed poor reliability (ICCs: 0.03 to 0.33). ICCs could not be calculated for three proscribed items (they occurred rarely or not at all). Internal consistency for the overall scale was measured using Kuder-Richardson's alpha (for dichotomous scoring); the coefficient was 0.94, indicating acceptable internal consistency.</p> <p><i>Construct validity:</i> Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) provided support for construct validity. An eight-item two-factor structure (factor loadings <math>&gt;\pm 0.40</math>) accounted for 62% of the variance. The two factors overlapped with the main components of BNI: i) discussing with patients their perspectives on their alcohol use, and ii) identifying their motives and plans for changing their drinking. A range of goodness-of-fit indices were used to determine the model fit. Both factors met the thresholds for the fit indices, excepting the chi-square test.</p>

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
5. BAS continued					<i>Discriminant validity:</i> Chi-square analyses compared percent occurrence of BAS items between BNI and DI; almost all BNI consistent items occurred in significantly higher proportions of BNI than DI.
6. BECCI (Lane et al., 2005)	Behaviour change consultations in healthcare settings	BCC	Items generated from a literature review. Item reduction and refinement based on expert consultation and scale piloting.	183 training videos and simulated consultations	<p><i>Reliability:</i> Single measure ICCs for two groups of consultations were 0.73 (smoking cessation) and 0.93 (diabetes), showing good to excellent rater reliability. Test-retest reliability (also good to excellent) was demonstrated using the same method. Internal consistency was assessed using Cronbach's alpha. The coefficients were 0.71 (for baseline consultations, before training), and 0.63 (for final consultations, after training); internal consistency was a little low for the latter (<math>\alpha &lt; 0.7</math>).</p> <p><i>Construct, content, and face validity:</i> Support for validity was provided during scale development; specifically, through expert consultation, construct explication (describing the relationship between item behaviours and abstract constructs), and scale piloting. The items developed were centrally related to BCC and focused on therapist behaviours.</p> <p><i>Sensitivity to change:</i> Change in BECCI scores before and after training was assessed using the standardised response mean (mean change in scale score divided by standard deviation of that change); the value was 1.76, showing high sensitivity to change (<math>&gt;0.8</math>).</p>

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
7. CBT for PTSD Fidelity Scale (Lu et al., 2012)	Serious mental illness and post-traumatic stress disorder (PTSD)	Cognitive-behavioural therapy (CBT) for PTSD program	Adapted from the Problem-Solving Treatment of depression for Primary Care (Hegel et al., 2004).	30 sessions	<i>Reliability:</i> The ICC for the individual items was 0.82, indicating excellent reliability among the five raters' scores. Cronbach's alpha coefficients for the early sessions of the CBT for PTSD program (sessions 1-3) and the later sessions (sessions 4-16) were 0.84 and 0.90, indicating acceptable internal consistency.
8. CBT Therapist Checklist	-	-	Derived from cognitive-behavioural therapy strategies and interventions.	-	N/A: Measure identified from contacting author, referenced in the 'Cognitive-Behavioral Coping Skills Treatment for Cocaine Dependence' manual (Carroll, 1997).
9. CE Therapist Rating Form	-	-	Derived from the strategies and interventions of compliance enhancement (CE).	-	N/A: Measure identified from contacting author, referenced in the CE manual (Carroll et al., 1999).
10. CM Clinician Rating Form	-	-	Informed by contingency management (CM) techniques.	-	N/A: Measure identified from contacting author, referenced in CM manual (Petty and Stitzer, 2002).
11. CPPS (Hilsenroth et al., 2005)	Outpatients receiving psycho-dynamic (PD), and non-PD therapies	PD therapies, and non-PD therapies	Based on reviews of the psychotherapy process literature (Blagys and Hilsenroth, 2000; 2002).	124 video recordings (105 PD, and 19 non-PD)	<i>Reliability:</i> ICCs (2,1) for CPPS PI and CPPS CB subscales were 0.93 and 0.95. ICCs for the 20 individual items ranged from 0.67 to 0.95, indicating good to excellent rater agreement. Cronbach's alpha for the two subscales was 0.92 and 0.94, showing high internal consistency.

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
11. CPPS continued					<p><i>Convergent validity:</i> The relationships between CPPS PI and subscales of two therapist activity measures evaluating similar forms of psychodynamic or supportive-expressive psychotherapy were examined using Pearson's correlation coefficient. Large effect sizes (<math>&gt;0.50</math>) were mostly found providing evidence of convergent validity; moderate correlations (<math>&gt;0.30</math>) were found between CPPS PI with adherence and competence of supportive techniques.</p> <p><i>Discriminant validity:</i> Seventeen non-PD sessions were matched with 17 PD sessions. An analysis of variance (ANOVA) compared the mean item and subscale scores between the two groups. PD therapy sessions had significantly higher CPPS PI item and subscale scores than the non-PD sessions. Non-PD sessions had significantly higher CPPS CB item and subscale scores than the PD therapy sessions, excepting one item for which there was no statistically significant difference.</p>
12. CSPRS-6 (Hill et al., 1992)	Psychiatric outpatients with a major depressive disorder	CBT, interpersonal therapy (IPT), and clinical management (ClinM)	Derived from treatment manuals, associated literature and therapy trainer consultation. Scale revisions based on psychometric testing.	720 audio recordings	<p><i>Reliability:</i> ICCs (3,8) for the three modality-specific scales and the two non-modality-specific scales ranged from 0.58 to 0.92; values indicated good to excellent consistency between the two raters' scores (eight raters in total, rotating pairs of two raters) for all but one non-modality-specific scale (Facilitative Conditions). Cronbach's alpha for the scales was calculated using the combined scores for the two raters, and ranged from 0.50 to 0.86; there was acceptable internal consistency, excepting two scales: ClinM (<math>\alpha=0.69</math>), and Explicit Directiveness (<math>\alpha=0.50</math>).</p> <p><i>Construct validity:</i> A principle components analysis (PCA) with a five factor solution accounted for 31% of the total variance; the structure underlying the ratings on the 96 items was not well represented by five components.</p>

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
12. CSPRS-6 continued					<i>Discriminant validity:</i> A multiple-groups profile analysis (similar to a multi-factor repeated measures ANOVA) showed an effect of treatment modality; subsequent ANOVAs were used to compare the subscales across therapies. CBT therapists scored higher on the CBT scale, IPT therapists scored higher on the IPT scale, and ClinM therapists scored significantly higher on the ClinM scale than the other two treatments.
13. CTACS (Barber et al., 2003)	Primary diagnosis of cocaine dependence	CT, SET, and IDC	Derived from the Collaborative Study Psychotherapy Rating Scale (CSPRS; Hollon et al., 1988), Cognitive Therapy Scale (CTS; Young and Beck, 1980), and CT manuals. Scale revisions based on piloting and reliability analyses.	134 audio recordings (92 CT, 20 SET, and 22 IDC)	<i>Reliability:</i> ICCs (2,2) for the 25 individual items among CT cases, separated for adherence and competence, ranged from 0.33 to 0.92. Inter-rater reliability was poor (<0.40) for four items, scale revised to exclude these items. ICCs for the 21-item total scale (among CT cases) for adherence and competence were 0.67 and 0.73, indicating good reliability among the two raters' scores. Cronbach's alphas for adherence and competence were 0.92 and 0.93, indicating high internal consistency.  <i>Discriminant validity:</i> Contrast analyses were conducted comparing CT to SE and IDC. The analyses showed significant differences ( $p < 0.005$ ) between CT and the other two treatments.
14. CTS-R (Blackburn et al., 2001a)	Patients receiving CT for a range of problems, e.g., social phobia.	CT	Based on the CTS (Young and Beck, 1980), and expert consultation.	102 video recordings	<i>Reliability:</i> ICCs (two-way model) for the total scores and individual items on the 13-item version (12-item version not validated) ranged from -0.14 <sup>a</sup> to 0.86, indicating poor to excellent reliability among the raters' scores (four raters paired with each other). Cronbach's alpha for each rater on the 13-item scale were 0.92, 0.95, 0.97, and 0.95; the values show very high internal consistency, indicating some redundancy in the scale (Bland, 2017).  <i>Face validity:</i> Face validity demonstrated through consultation with the raters.

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
14. CTS-R continued					<i>Sensitivity to change:</i> Eleven trainee CT therapists saw two clients at different stages of training. A paired t-test compared the mean total scores for the first and second client on the 13-item version, a statistically significant difference was found. Changes in competence on each item were also examined; statistically significant differences were found for seven items.
15. GROMIT	-	-	-	-	N/A: Measure identified from contacting author.
16. IAC Treatment Fidelity Instrument (Torrey, 2012)	Student nurses portraying patients as substance using pregnant women.	I Am Concerned (IAC) brief opportunistic intervention	Derived from the IAC treatment manual. Scoring guided by the literature, and YACSII <sup>(No.26)</sup> . Scale revisions based on expert consultation and piloting.	49 audio recordings of simulated clinic sessions	<i>Reliability:</i> ICCs (model not specified) for adherence and competence were 0.64 and 0.62. ICCs for the 18 individual items, separated for adherence and competence, ranged from -0.07 to 0.81. For 13 items, there was fair to excellent consistency between the three raters' scores. Four low-ICC adherence items were associated with high levels of percentage agreement and low data variance (" <i>ratings fall under only a few scale scores</i> ", p.814). Cronbach's alpha for the scale was 0.72, indicating acceptable internal consistency. Internal consistency coefficients for adherence and competence were 0.54 and 0.56. The Spearman-Brown adjustment (a test to estimate the reliability of a measure after changing the length) estimated the coefficient values if the number of items were doubled, the coefficients increased to 0.84 for the scale, and 0.70 and 0.72 for the subscales.  <i>Content validity:</i> Support for content validity was provided during the development of the measure. Two experts were consulted on the clarity, sufficiency, and relevance of the items. A content validity index (CVI) was computed to quantify the extent of agreement between the experts. A CVI of 1.0 was achieved, indicating complete agreement.



Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
17. ITRS (Martino et al., 2008)	Outpatients seeking substance use treatment	Motivational enhancement therapy (MET), and counselling as usual (CAU)	Adapted from the Yale Adherence and Competence Scale (YACS; Carroll et al., 2000).	425 recordings (206 MET, and 219 CAU)	<p>Thirty (of 42) items validated: interventions consistent with motivational interviewing (MI-consistent; <math>n=10</math>), MI-inconsistent interventions (<math>n=10</math>), and general therapist interventions (<math>n=10</math>).</p> <p><i>Reliability:</i> ICCs (3,1) for the 30 individual items, separated for adherence and competence, ranged from 0.55 to 0.99. For 28 items, there was good to excellent rater consistency (15 raters each scored 15 recordings); two items showed fair reliability. Internal consistency analyses centred on MI-inconsistent and general therapists interventions. MI-inconsistent items occurred infrequently; when present, three items occurred more often. Cronbach's alpha for these three items was 0.84. Of the general therapist interventions, two items frequently occurred. Cronbach's alpha for the two items was 0.39, indicating poor internal consistency; these items were therefore excluded from subsequent analyses.</p> <p><i>Construct validity:</i> Construct validity was examined through a CFA. MI-consistent adherence items converged to form two <i>a priori</i> defined factors related to the core components of MI: fundamental MI skills, and advanced MI skills. A range of goodness-of-fit indices were used to determine the model fit. Both factors met most thresholds for the fit indices, suggesting an acceptable model fit. Advanced MI skills fitted best with session 2 and session 3 (therapists delivered three sessions of MET or CAU).</p>
18. MACT Rating Scale (Davidson et al., 2004)	Adults with recurrent episodes of self-harm	Manual assisted cognitive therapy (MACT)	Based on the MACT treatment manual.	49 audio recordings	<i>Reliability:</i> An ICC of 0.66 suggested good reliability among two raters' scores for 12 tapes.

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
19. MI-CTS (Haddock et al., 2012)	Adults with psychosis and a substance use problem.	Integrated motivational interviewing and cognitive behavioural therapy (MI-CBT)	Derived from existing CBT and MI fidelity measures, and discussion with practicing clinicians. Revisions based on scale piloting.	34 audio recordings	<p>Sixteen (of 19) items validated: Section A (<math>n=5</math>), and Section B (<math>n=11</math>).</p> <p><i>Reliability:</i> Percentage agreement (adherence versus non-adherence) for the two subscales was 100% when a “<i>criterion of within two points of each other was taken</i>” (p.42; four raters each scored 10 tapes). Percentage agreement for the individual items ranged from 55.56% to 100%.</p> <p><i>Convergent validity:</i> Ratings on the MI-CTS were compared with scores on the Cognitive Therapy Scale for Psychosis (CTS-Psy; Haddock et al., 2001) and the Motivational Interviewing Treatment Integrity Scale (MITI; Moyers et al., 2005) using Spearman’s rho. Support for convergent validity was found through statistically significant correlations between: i) Section A and the overall empathy rating on the MITI (<math>\rho = 0.52</math>), and ii) Section B and the CTS-Psy total scores (<math>\rho = 0.46</math>) and subscale scores (<math>\rho = 0.46</math> and <math>0.43</math>).</p>
20. MISTS <sup>3</sup> (Madson et al., 2005)	Substance use	MET	Based on the CSPRS (Hollon et al., 1988), YACS (Carroll et al., 2000) and MITI (Moyers et al., 2005)	50 audio recordings	<p><i>Reliability:</i> The overall generalisability coefficient (akin to an ICC; multiple sources of error are represented in a single analysis and are derived from ANOVA models) between the two raters was high (<math>\rho^2 = 0.79</math>). ICCs (model not specified) for the 16 items ranged from 0.41 to 0.81, indicating fair to excellent reliability among the two raters’ scores.</p> <p><i>Convergent and discriminant validity:</i> Assessed by comparing the MISTS total score with the YACS six subscale scores using Pearson’s correlation coefficient. Support for validity was found though correlations between the MISTS total score and four of the YACS subscales; two subscales for convergent (<math>r = 0.72</math> and <math>0.70</math>, statistically significant), and two subscales for discriminant (<math>r = 0.15</math> and <math>-0.27</math>, not statistically significant).</p>

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
21. MTRS (DeRubeis et al., 1982)	Outpatients with a major depressive disorder	CBT, and IPT	Derived from the CTS (Young and Beck, 1980), the Global Rating Form and Process Rating Form (O'Malley et al., 1988), and CT and IPT treatment manuals.	12 tapes in audio and video format (6 CBT, and 6 IPT)	<p><i>Reliability:</i> Pearson's correlation coefficient for the 48 items was used to examine inter-rater reliability among six pairs of raters; values ranged from -0.29 to 0.92, scores for nine items showed a weak association (<math>r &lt; 0.30</math>).</p> <p><i>Construct validity:</i> Items were subjected to a PCA, nine factors were found, accounting for 85% of the total variance. Factors 5 to 9 were disregarded (e.g., included only 1 item). The remaining 4 factors accounted for 69% of the variance: CBT techniques (<math>n=15</math>), general therapeutic skills (<math>n=10</math>), therapist directiveness (<math>n=4</math>), IPT techniques (<math>n=3</math>). A direct discriminant function analysis correctly classified the two treatments in all 12 tapes based on the factors: CBT and IPT techniques. Scale revised to 32 items.</p> <p><i>Discriminant validity:</i> t-tests compared item scores across the two therapies; 34 (of 48) items discriminated between the therapies (<math>p &lt; 0.05</math>).</p>
22. SPRS (Shapiro and Startup, 1992)	Adults with a major depressive disorder	Exploratory therapy (ET), and prescriptive therapy (PT)	Based on the CSPRS (Hollon et al., 1988), the coding scheme developed by Goldberg et al. (1984), and expert consultation.	220 audio recordings (108 ET, and 112 PT)	<p><i>Reliability:</i> ICCs (1,1) for the therapy subscales (ET and PT) were 0.78 and 0.85, indicating excellent rater agreement. The subscale, facilitative conditions (FC), was reported to have low reliability. Because of this an ICC (1,2) was computed; the coefficient was 0.65, indicating good reliability. ICCs for the 27 individual items on the ET and FC subscales ranged from -0.11 to 0.80, indicating variable reliability for the two raters across both therapies. Cronbach's alpha was computed for the ET and FC subscales, using the average of the two raters' scores for ET sessions only; the values were 0.85 and 0.81, indicating acceptable internal consistency.</p> <p><i>Construct validity:</i> Examined through EFA using ET sessions. A 20-item, three-factor solution accounted for 39% of total variance (factor loadings <math>&gt; 0.30</math>). The factors (emotional experiencing, non-specific techniques, enhancing insight) support ETs integration of psychodynamic and experiential constructs. Empirical subscales were constructed summing items with loadings <math>&gt; 0.30</math> on the emotional experiencing and enhancing insight factors; Cronbach's alpha for these subscales were 0.81 and 0.74.</p>

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
23. TPRS (Karno and Longabaugh, 2007)	Adults with a diagnosis of alcohol abuse or alcohol dependence	CBT, MET, and twelve-step facilitation (TSF)	Not specified.	15 to 25 minute segments of 548 video recordings	<i>Reliability:</i> The emotion-focus subscale (4-items) showed excellent inter-rater reliability (ICC = 0.89) and acceptable internal consistency ( $\alpha = 0.90$ ). The confrontation (3-items) and structure (3-items) subscales were combined as a measure of therapist directiveness. The combined scale showed excellent inter-rater reliability (ICC = 0.75). The article references additional research to highlight that confrontation and structure were separate factors, each with good internal consistency (factor loadings >0.7).
24. TSF-ACES (Campbell et al., 2013)	Outpatients receiving treatment for stimulant abuse	Individual (and group) TSF integrated into treatment as usual	Based on adherence scales developed for the STAGE-12 trial (Stimulant Abuser Groups to Engage in 12-Step; Donovan et al., 2013); the scales were adapted from previous TSF fidelity measures.	966 audio recordings (476 individual TSF sessions)	<i>Reliability:</i> ICCs (“a two-way random mixed model”, p.173) for the single-item summary measures (global empathy and global session ratings) were 0.69 and 0.80. ICCs for adherence, competence, and proscribed behaviours (based on means of multiple-items) were 0.91, 0.90, and 0.83. The coefficients indicated good to excellent inter-rater reliability; 59 sessions double rated.  Overall, internal consistency was acceptable; weighted mean alpha coefficients for adherence and competence across session types were 0.69 and 0.71. Weighted means were used because TSF-ACES comprised four session-specific rating forms (one for groups and three for individual sessions 1-3), items in the individual forms varied depending on whether clients attended 12-step meetings the previous week. Only one individual form (session 3 with 12-step meeting attendance) showed acceptable internal consistency for adherence and competence ( $\alpha = 0.81$ and 0.84).  <i>Convergent validity:</i> The relationships between average TSF-ACES summary measure scores and Helping Alliance questionnaire-II (HAq-II) scores were examined using Pearson’s correlation coefficient. Support for convergent validity was found through statistically significant correlations between HAq-II (scores at treatment end) and adherence ( $r = 0.31$ ), competence ( $r = 0.28$ ), and global session rating ( $r = 0.21$ ).

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
25. UKATT PRS (Tober et al., 2008)	Clients accessing treatment of alcohol problems	MET and social behaviour and network therapy (SBNT)	Adapted from the MATCH (Matching Alcoholism Treatment to Client Heterogeneity) Tape Rater Scale (Carroll et al., 1998a)	452 video recordings (259 MET, and 193 SBNT)	<p><i>Reliability:</i> ICCs (3,2) for the 12 MET and 8 SBNT items, rated for frequency, ranged from 0.28 to 0.96. Most items showed good to excellent levels of consistency between the two raters' scores. Two MET items showed fair reliability, and one SBNT item showed poor reliability. Cronbach's alpha coefficients for frequency of MET and SBNT items ranged from 0.65 to 0.76. Internal consistency was a little low for 5 MET items and 1 SBNT item.</p> <p><i>Construct validity:</i> The factor structure of the scale was examined by PCA using 20 treatment-specific items. A one-factor solution was found, accounting for 26% of the variance: "all MET items had positive loadings and all SBNT items had negative loadings suggesting a treatment component where the more MET was practiced, the less SBNT was practiced" (p.677).</p> <p><i>Convergent validity:</i> Examined by comparing quality ratings for MET and SBNT with global ratings of individual therapists' skill (low, medium, or high); ratings were made by the treatment-specific supervisors. An ANOVA showed a statistically significant differences (p&lt;0.001) in quality ratings for the two treatments across the three (low, medium, high) categories.</p> <p><i>Discriminant validity:</i> t-tests were used to compare the mean scores for item frequency between MET and SBNT. There were significant differences between item scores across the treatments, with higher scores on MET items for MET sessions and higher scores on SBNT items for SBNT sessions.</p>

Measure (Article Author, Year)	Population for Validation	Therapies Rated for Validation	Development of measure	Validation Data	Main Validation Analyses
26. YACS <sup>4</sup> (Carroll et al., 2000)	Adults with cocaine dependence and concurrent alcohol abuse or dependence	TSF, ClinM, and CBT	Adapted from five previous fidelity measures, including the ITRS <sup>(No.17)</sup> . Items based on TSF, ClinM, and CBT treatment manuals.	576 recordings	<p><i>Reliability:</i> ICCs (2,1) for the six subscales (three treatment-specific and three non-treatment-specific), separated for adherence and competence, ranged from 0.71 to 0.95, indicating good to excellent rater reliability (five raters each scored 19 recordings). ICCs for the 40 individual items, separated for adherence and competence, ranged from 0.06 to 0.84, indicating variable reliability.</p> <p><i>Convergent validity:</i> Examined by comparing the YACS subscales with four alliance measures using Pearson's correlation coefficient. Support for convergent validity was found through significant correlations between the YACS general support subscale, separated for adherence and competence, and each of the alliance measures (<math>r = 0.28</math> to <math>0.57</math>). The TSF subscale scores also showed significant correlations with three of the four alliance measures (<math>r = 0.25</math> to <math>0.43</math>).</p> <p><i>Construct validity:</i> Examined through a CFA using adherence ratings from 83 recordings (early treatment sessions). It was hypothesised that each item would be associated with a specific subscale. Five subscales met thresholds for the four goodness-of-fit indices. The cognitive behaviour treatment subscale satisfied one of the indices, results were a little low for the remaining indicators.</p> <p><i>Discriminant validity:</i> Evaluated in two ways: i) by comparing mean subscale scores across treatments, and ii) by performing a multiple-groups profile analysis. The treatment-specific subscale scores were significantly different across the three treatments: TSF subscale scores were significantly higher for TSF sessions, ClinM subscale scores were significantly higher for ClinM sessions, and CBT scores were significantly higher for CBT sessions compared to the other two treatments. The profile analysis provided additional support for the findings; the subscales were able to differentiate between the three treatments.</p>

<sup>1</sup> Cohen's *d* (observed difference divided by the pooled standard deviation) effect size indices: 0.8 is a large effect, 0.5 a moderate effect, and 0.2 a small effect (Bland, 2017).

<sup>3</sup> Madson et al. (2005) validated the 16-item MISTS, not the 20-item MISTS Revised (validation of MISTS Revised not found).

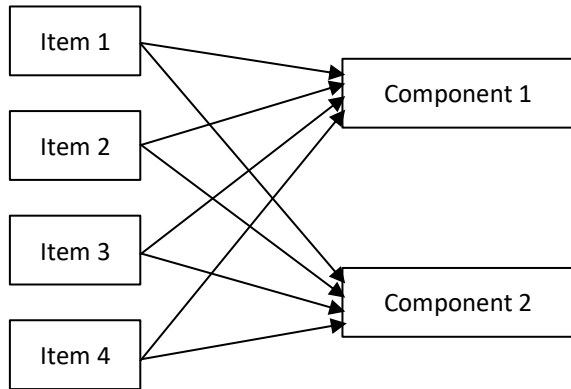
<sup>4</sup> Carroll et al. (2000) validated the 55-item YACS, not the 68-item YACSII (validation of YACSII not found).

<sup>a</sup> Negative values of the ICC have no theoretical legitimacy, and indicate poor agreement (Giraudeau, 1996). Negative ICCs can occur when the rater variability (within-groups) exceeds the therapist (or subject) variability (between-groups).

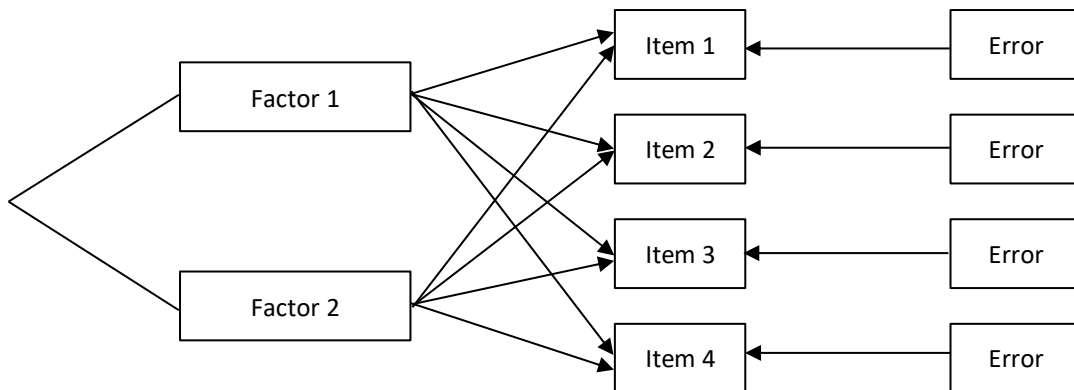
**ACS-IDCCD** = Adherence/Competence Scale for Individual Drug Counseling for Cocaine Dependence; **ACS-SEC** = Adherence/Competence Scale for Supportive-Expressive Therapy for Cocaine Dependence; **ADAPTA PRS** = ADAPTA (Addressing Drinking Among Patients: comparing Two Approaches) Process Rating Scale; **AESOPS PRS** = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale; **BAS** = Brief Negotiation Interview Adherence Scale; **BECCI** = Behaviour Change Counselling Index; **CBT for PTSD Fidelity Scale** = Cognitive-Behavioral Therapy for Post-Traumatic Stress Disorder Fidelity Scale; **CBT Therapist Checklist** = Cognitive Behavioral Therapy Therapist Checklist; **CE Therapist Rating Form** = Compliance Enhancement Therapist Adherence/Competence Rating Form; **CM Clinician Rating Form** = Contingency Management Clinician Adherence/Competence Rating Form; **CPPS** = Comparative Psychotherapy Process Scale; **CSPRS-6** = Collaborative Study Psychotherapy Rating Scale – Form 6; **CTACS** = Cognitive Therapy Adherence and Competence Scale; **CTS-R** = Cognitive Therapy Scale – Revised; **GROMIT** = Global Rating of Motivational Interviewing Therapist; **IAC Treatment Fidelity Instrument** = I Am Concerned Treatment Fidelity Instrument; **ITRS** = Independent Tape Rater Scale; **MACT Rating Scale** = Manual Assisted Cognitive Therapy Rating Scale; **MI-CTS** = Integrated Motivational Interviewing and Cognitive Behavioural Therapy Fidelity Scale; **MISTS Revised** = Motivational Interviewing Supervision and Training Scale Revised; **MTRS** = Minnesota Therapy Rating Scale; **SPRS** = Sheffield Psychotherapy Rating Scale; **TPRS** = Therapy Process Rating Scale; **TSF-ACES** = Twelve Step Facilitation Adherence Competence Empathy Scale; **UKATT PRS** = UK Alcohol Treatment Trial Process Rating Scale; **YACSII** = Yale Adherence and Competence Scale Second Edition.

### B.5 Methods for establishing construct validity

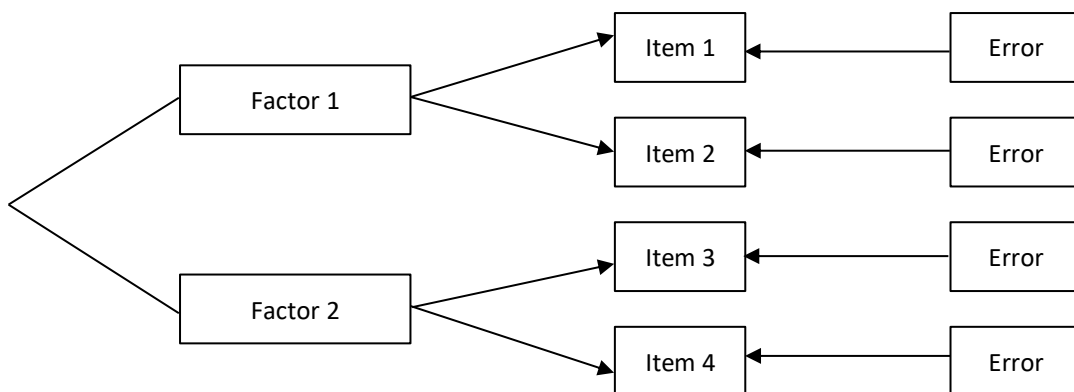
The figures provided below give a visual representation of the three methods used in the articles to examine construct validity of the measures: principal components analysis (PCA), exploratory factor analysis (EFA), and confirmatory factor analysis (CFA).



**PCA model diagram**



**EFA model diagram**



**CFA model diagram**



## B.6 Variants of the intraclass correlation coefficient

Model; Abbr.	Model Overview	Study Design	Model variants; Abbr.
One-way random-effects; ICC(1)	<ul style="list-style-type: none"> <li>Does not separate the effect of therapist and rater (one-way model).</li> <li>Assumes both therapists and raters were randomly selected (random effects).</li> <li>Usually gives the lowest ICCs.</li> </ul>	Recordings scored by different raters.	<ul style="list-style-type: none"> <li>Single measures, absolute agreement; ICC(1,1) agreement</li> <li>Average measures, absolute agreement; ICC(1,k) agreement</li> </ul>
Two-way random-effects; ICC(2)	<ul style="list-style-type: none"> <li>Separates the effect of therapist and rater (two-way model).</li> <li>Assumes both therapists and raters were randomly selected (random-effects).</li> <li>Findings can be generalised.</li> </ul>	Recordings scored by the same raters. Raters regarded as being randomly selected.	<ul style="list-style-type: none"> <li>Single measures, consistency; ICC(2,1) consistency</li> <li>Single measures, absolute agreement; ICC(2,1) agreement</li> <li>Average measures, consistency; ICC(2,k) consistency</li> <li>Average measures, absolute agreement; ICC(2,k) agreement</li> </ul>
Two-way mixed-effects; ICC(3)	<ul style="list-style-type: none"> <li>Separates the effect of therapist and rater (two-way model).</li> <li>Assumes the raters were not randomly selected: therapists are a random effect and raters a fixed effect (mixed effects).</li> </ul>	Recordings scored by the same raters. Raters not randomly selected.	<ul style="list-style-type: none"> <li>Single measures, consistency; ICC(3,1) consistency</li> <li>Single measures, absolute agreement; ICC(3,1) agreement</li> <li>Average measures, consistency; ICC(3,k) consistency</li> <li>Average measures, absolute agreement; ICC(3,k) agreement</li> </ul>

Abbr. = Abbreviation; ICC = intraclass correlation coefficient.

Model variants based on definitions provided by Shrout and Fleiss (1979) and McGraw and Wong (1996); Single measures = Reliability estimates based on a single rater (Trevethan, 2017); Average measures = Reliability estimates based on the mean of multiple raters (Koo and Li, 2016); Absolute agreement = the degree to which different raters assign the same scores (rater variability contributes to the denominator of the ICC) (Streiner et al., 2015); Consistency = the degree to which different raters' scores are correlated (rater variability does not contribute to the denominator of the ICC) (Landers, 2015).

## **Appendix C**

### **Chapter 4 Appendices**

Appendix C focuses on the appendices for chapter 4, generating an item pool. Appendix C contains one table:

- i) Stage 2 refinements to the thematic structure made by the researcher and the supervision team (p.241).

### C.1 Refinements to the thematic structure made by the researcher and the supervision team

Meta-theme	Theme	Sub-theme	Decision	Revised Theme	Rationale
Session management	Focus and structure	Maintaining Structure	Revised	Problem focused	Items lacked specificity, or indicated how therapists might maintain session structure, overlapping with other 'focus and structure' subthemes. Items captured by the new theme 'problem focused'.
		Agenda setting	Revised		Items focused on therapists setting and following an agenda, captured by 'problem focused'. Complex items, combining agenda setting and collaboration, were captured by 'problem focused' and 'collaboration' (meta-theme 5, core skills).
		Consistency of problem focus	Revised		Items related to therapists keeping the session focused on relevant topics, captured by 'problem focused'.
		Directiveness	Revised		Items focused on the level of direction or guidance given by therapists during the session, captured by 'problem focused' and meta theme 5 (core skills), e.g., 'collaboration'.
		Time management	Revised		Items were specific and limited in number ( $n=4$ ), captured indirectly by 'problem focused'.
	Therapy rationale		Removed		Most items were therapy specific. Items that did not specify a therapeutic approach overlapped with the subtheme 'agenda setting', and were therefore captured by 'problem focused'.
	Session content		Removed		Items were therapy specific, focusing on the type of therapy or techniques delivered during the session.

Meta-theme	Theme	Sub-theme	Decision	Revised Theme	Rationale
Session management	Summarising		Removed		Items related to therapists use of summaries during the session. Items overlapped with 'maintaining structure', which was captured by 'problem focused', and meta-theme 5 (core skills), a form of reflective listening.
	Termination of therapy		Removed		Therapy session specific (later sessions), focusing on the termination of therapy.
Medication and case management	Medication		Removed		Therapy and client specific, focusing on the client's medication.
	Case management		Removed		Items were concerned with therapists encouraging, facilitating, and reviewing the client's use of self-help groups and other services. Captured indirectly by other themes, e.g., the client may agree to attend a self-help group between therapy sessions; this example would be captured by 'homework assigned', meta-theme 4 (interventions to increase awareness).
Interventions to increase awareness	Explores behaviours	Explores general behaviours	Revised	Exploring behaviours	Sub-theme retained, minor revision to name made. Items focused on therapists exploring clients' problem behaviours, without reference to substance use or other concerns.
		Explores substance use behaviours	Revised	Exploring substance use	Sub-theme retained, minor revision to name made. Items focused on the assessment and monitoring of clients' substance use problems.
	Explores psychological health		Removed		Items related to therapists' assessment of the client's psychological health, captured by 'explores behaviours'.
	Explores general functioning		Removed		Items focused on therapists' assessment of the client's general level of functioning, captured by 'exploring behaviours'.

Meta-theme	Theme	Sub-theme	Decision	Revised Theme	Rationale
Interventions to increase awareness	Explores future		Removed		One item was concerned with the client's future, captured by 'exploring behaviours'.
	Ambivalence		Revised	Exploring impact of substance use  Exploring pros and cons of change	Items concerned therapists' exploring the client's conflicting thoughts and feelings about changing their behaviour. While most items did not explain the term ambivalence ( $n=5$ ), two asked multiple questions, indicating how ambivalence might be explored. Asking multiple questions within the same item reduces item reliability. Therefore, ambivalence was separated into two themes: i) 'exploring impact of substance use', therapists' exploring the impact of the client's substance use, and ii) 'exploring pros and cons of change', therapists increasing the client's awareness of what might be gained and lost by changing their substance use.
	Understanding problems and behaviours	Understanding behaviours	Removed		Items focused on therapists helping the client to gain an understanding of their problem behaviours. No obvious exemplars; a range of issues were covered. Items captured by 'exploring behaviours' and 'exploring impact of substance use'.
		Understanding relationships	Removed		Items related to therapists helping the client to understand their interpersonal relationships. No obvious exemplars; a range of issues covered. Most items coded from the psychotherapy measures <sup>1</sup> , suggesting limited relevance to BATS. Items captured by 'exploring impact of substance use' and 'complex reflections' (meta-theme 5, core skills).
		Understanding thoughts and feelings	Removed		Items focused on therapists increasing the client's awareness of their thoughts and feelings. No obvious exemplars; a range of issues were covered. Most items coded from the psychotherapy measures <sup>2</sup> , suggesting limited relevance to BATS. Items captured by 'exploring impact of substance use' and 'complex reflections' (meta-theme 5, core skills).

Meta-theme	Theme	Sub-theme	Decision	Revised Theme	Rationale
Interventions to increase awareness	Resistance		Removed		Items related to therapists addressing the client's resistance to change their behaviour and engage in treatment. Captured indirectly by 'complex reflections' (meta-theme 5, core skills) and 'exploring pros and cons of change'.
	Wishes, dreams and childhood		Removed		Items were concerned with the client's wishes, dreams, and/or childhood experiences. Such techniques are not usually used in the treatment of substance use problems, e.g., focusing on unconscious wishes and early childhood memories. Indeed, most items were coded from the psychotherapy measures <sup>3</sup> , suggesting limited relevance to BATS.
	Therapy specific techniques		Removed		Items were therapy specific.
Interventions to change behaviour	Advice		Removed		Items were concerned with therapists' giving advice to the client, captured indirectly by other themes, e.g., therapists may offer advice when assigning homework.
	Behaviour change	Moving towards change	Revised	Talking about change	Sub-theme retained, minor changes to name made. Items focused on therapists encouraging the client to talk about behaviour change.
		Commitment to change	Removed		Items were concerned with therapists encouraging the client to make a commitment to change their substance use, captured by 'talking about change', 'treatment goals' and 'behaviour change planning'.
		Making a plan	Revised	Behaviour change planning	Items focused on therapists developing a plan with clients for changing and achieving treatment goals. Some items were complex, covering different aspects of change planning. Asking multiple questions within the same item can reduce item reliability, therefore, 'making a plan' was separated into three sub-themes: i) 'behaviour change planning', focusing on developing a plan, ii) 'identifying sources of support', identifying supportive people available to help the client, and iii) 'involving others', discussing the nature of that support.
			Identifying sources of support		
				Involving others	

Meta-theme	Theme	Sub-theme	Decision	Revised Theme	Rationale
Interventions to change behaviour	Behaviour change	Treatment goals	Retained	Treatment goals	Sub-theme retained. Items were relevant, focusing on therapists discussing the client's goals for treatment.
	Homework		Revised	Homework assigned Homework reviewed	Items were concerned with therapists: i) assigning tasks for the client to do between sessions, and ii) reviewing previously assigned tasks. To improve item reliability, 'homework' was separated into two themes: 'homework assigned', and 'homework reviewed'.
	Providing information		Removed		Items focused on therapists providing clients or family members with information, captured indirectly by other items, e.g., therapists may provide information when assessing the client's substance use.
	Developing skills		Removed		Items related to therapists developing the client's skills, considered therapy session specific, with the techniques not universally or consistently used.
	Restructuring thoughts		Removed		Items were concerned with changing the client's thoughts and beliefs, captured by 'behaviour change planning', and the revised homework themes.
	Therapy specific techniques		Removed		Items were therapy specific.
Core skills	Supporting client self-efficacy		Revised	Strengths and affirmations	Theme retained, minor changes to name made. Items focused on therapists' recognising and reinforcing the client's strengths, abilities, or efforts to change.
	Collaboration		Retained	Collaboration	Theme retained, items focused on the therapist and the client working together.

Meta-theme	Theme	Sub-theme	Decision	Revised Theme	Rationale
Core skills	Empathic relationship	Empathy	Retained	Empathy	Sub-theme retained. Items focused on therapists being able to respond empathetically to the client. 'Empathy' overlapped with the other 'empathic relationship' subthemes.
		Acceptance, respect, and being non-judgemental	Removed		Items were concerned with therapists conveying respect, and appearing non-judgemental and accepting of what the client said, captured by 'empathy'.
		Warmth and genuineness	Removed		Items focused on therapists conveying warmth and genuineness, captured by 'empathy'.
		Support and reassurance	Removed		Items related to therapists being supportive and reassuring to the client, captured by 'empathy'.
		Being attuned and attentive	Removed		Items concentrated on therapists being attuned and attentive to what the client said, and were captured by 'empathy'.
		Rapport	Removed		Items focused on how well the therapist and client got along, captured by 'empathy'.
		Sensitivity and concern	Removed		Items looked at whether therapists expressed concern and were sensitive to the client's perspective, and were captured by 'empathy'.
	Presentation	Removed		Items referred to therapists' style of presentation, captured indirectly by other items, e.g., the expressiveness of a therapists voice overlapped with 'empathy'.	
Self-disclosure	Removed		Items focused on therapists' use of appropriate self-disclosure, a technique therapists may use to convey 'empathy' and foster 'rapport'.		



Meta-theme	Theme	Sub-theme	Decision	Revised Theme	Rationale
Core skills	Asking questions		Removed		Items were concerned with therapists asking questions, captured indirectly by other items, e.g., therapists will ask questions when exploring pros and cons of change (meta-theme 3, interventions to increase awareness).
	Reflective listening		Revised	Simple reflections Complex reflections	Items related to therapists' attempts to communicate understanding of what the client said. A distinction was made between simple and complex reflections. 'Reflective listening' was separated into two themes: i) simple reflections, therapists repeating or slightly rephrasing what the client had said, and ii) complex reflections, therapists adding new meaning to the client's comments.
	Overall performance		Removed		Items assessed therapists' overall performance by considering the appropriateness of the therapeutic techniques used, the level of skill shown, and the overall quality of the session. Items captured by the scale as a whole.
	Negative therapist attributes		Removed		Items referred to therapist attributes considered detrimental to therapeutic progress and behaviour change. Given the limited number of items needed for BATS, focus was given to prescribed behaviours.

<sup>1</sup> Forty eight items coded 'understanding relationships'; thirty six items coded from ACS-SEC, CPPS, CSPRS-6, and SPRS (Barber, 1997; Hilsenroth et al., 2005; SPR Project Staff, 1984; Shapiro and Startup, 1990).

<sup>2</sup> Fifty three items coded 'understanding thoughts and feelings'; forty four coded from ACS-SEC, CPPS, CSPRS-6, SPRS, and TPRS (Fisher et al., 2000).

<sup>3</sup> Nine items coded 'wishes dreams and childhood'; eight items coded from ACS-SEC, CPPS, CSPRS-6, SPRS, and TPRS.

**ACS-SEC** = Adherence/Competence Scale for Supportive-Expressive Therapy for Cocaine Dependence; **CPPS** = Comparative Psychotherapy Process Scale, **CSPRS-6** = Collaborative Study Psychotherapy Rating Scale – Form 6; **SPRS** = Sheffield Psychotherapy Rating Scale; **TPRS** = Therapy Process Rating Scale.

## **Appendix D**

### **Chapter 5 Appendices**

Appendix D focuses on the appendices for chapter 5, agreeing the content. The following appendices are included:

- i) A screenshot from the round one questionnaire of the participant information (p.249).
- ii) A screenshot from the round one questionnaire of the participant consent form (p.252).
- iii) An example screenshot of the main body of the round one questionnaire (p.253).
- iv) A table showing the participants grouped by demographic information to gauge individual participation across rounds (p.255).
- v) A statistical summary of participants' ratings from round one (p.256).
- vi) A statistical summary of participants' ratings from round two (p.257).
- vii) The BATS (p.258).

## D.1 Participant information screenshot from round one

# Delphi Study - Therapist Rating Scale Development (Round 1, Version 1)

0% complete

---

## Page 1: Participant information

### An invitation

I am inviting you to take part in a research study to develop a practical measure for evaluating addiction treatment delivery in routine practice. The study is being conducted at the University of Leeds. Please take some time to read the following information before deciding whether to take part.

### Study aims

The study is developing the Brief Therapist Rating Scale (BATS), a tool for evaluating therapist delivery of psychological therapies used for drug and alcohol use problems in routine practice. To do this, the study aims to draw a consensus on which items to include and how best to rate them from experts and experienced practitioners in the fields of addiction and psychotherapy.

BATS will be used to support training and routine supervision, which has the potential to impact on therapist competence and improve client outcomes. To have utility in routine practice, the BATS needs to be:

1. Transtheoretical, applicable to range of therapies
2. Capable of monitoring therapist competence
3. A brief and easy to use measure

### Why am I being approached?

You have been identified as a potential expert participant – you have expertise in the field of addiction and/or psychotherapy, and may be interested in taking part. You were identified by one of my supervisors and/or nominated by another expert participant.

### What will I need to do?

If you agree to take part, you will be asked to complete: i) a short demographic questionnaire, and ii) three rounds of questionnaires, using Bristol Online Surveys (BOS). The questionnaires will involve the following:

**Round 1** – you will be given a list of items. You will be asked to rate:

- The importance you place on each item for inclusion in the BATS.
- The comprehensibility of each item. Considering potential users of the BATS, how easy are the items to understand and to rate?

You will be given an opportunity to comment on individual items, suggest item rewording and other items you feel should be included.

**Rounds 2 and 3** – based on your feedback from the previous round, you will be asked to rate a revised list of items to agree on the content of the BATS. You will be given an opportunity to comment on individual items and the item pool in general, with emphasis on layout and scoring. Rounds 2 and 3 will ensure that we end up with a scale that best reflects all of our opinions.

### **Do I have to take part?**

No, it is up to you whether or not you decide to take part. You can withdraw your consent for each of the questionnaires before they are submitted, but once submitted data will be automatically anonymised and therefore cannot be withdrawn. If you choose to withdraw your consent, you do not have to give a reason and you will not incur any consequences from us. Please feel free to contact me using the details below if you have any questions or you would like more information.

### **Confidentiality and ethics**

The study has received ethical approval from the School of Medicine Research Ethics Committee (MREC16-008). All data will be treated as confidential. Data will be anonymised at the point of download. Your personal details will not be published. The research supervisory team will have access to the data to assist with analysis, write up, and dissemination.

### **How will my data be stored?**

Questionnaire data will be stored online using Bristol Online Surveys. Data will be extracted (e.g. to word or excel) for analysis purposes. Data will be stored on a password-protected area of the secure University of Leeds network. All data will be stored for a period of three years following the end of data collection and then disposed of securely.

### **What are the benefits of taking part?**

There is no direct benefit, for example, payment or reimbursement of expenses. The data we gather will be used to develop BATS. The scale has a number of planned applications that may be of potential benefit to you and colleagues working in the fields of addiction and psychotherapy:

1. Use in routine clinical supervision to inform feedback on clinical skills.
2. Use in evaluating the training of therapists.
3. Use in fidelity rating in research projects, evaluating whether therapies were delivered as intended.

BATS will be made publically available and the findings from the study shared.

### **What are the potential risks of taking part?**

There are no anticipated risks from taking part. Completing the questionnaires will involve around 20 minutes for the first round, slightly less for the second and third rounds.

### **What if there is a problem?**

You are welcome to raise any concerns about the study with me or my supervisors using the contact details below.

## Contact details

Helen Crosby ([umhfc@leeds.ac.uk](mailto:umhfc@leeds.ac.uk))

Academic Unit of Psychiatry and Behavioural Sciences

Leeds Institute of Health Sciences

University of Leeds

Charles Thackrah Building

101 Clarendon Road

Leeds

LS2 9LJ

Supervised by:

- Dr. Gary Latchford ([G.Latchford@leeds.ac.uk](mailto:G.Latchford@leeds.ac.uk)), and Dr. Bridgette Bewick ([B.M.Bewick@leeds.ac.uk](mailto:B.M.Bewick@leeds.ac.uk)); address as above.
- Dr. Gillian Tober ([gillian.tober@nhs.net](mailto:gillian.tober@nhs.net)); address as follows:

The Coach House

Leeds Addiction Unit Training and Research Department

19 Springfield Mount

Leeds

LS2 9NG.

## Useful services

1. Addaction (<http://www.addaction.org.uk/>)
2. Alcoholics Anonymous (tel: 0845 796 7555)
3. Drinkline (tel: 0300 123 1110)

## D.2 Consent page screenshot from round one

# Delphi Study - Therapist Rating Scale Development (Round 1, Version 1)

 8% complete

---

## Page 2: Consent form

Please take time to read the following statements:

- I have have read and understood the study information.
- I have had the opportunity to ask questions, and any questions I have asked have been answered satisfactorily.
- I understand that my participation is voluntary and that I can withdraw from the study.
- I understand what will happen to my data, how it will be stored, and who will have access to it.
- I agree that anonymised quotes from my completed questionnaires can be used in the write-up and dissemination of findings from the study.
- I agree to take part in the study.

If you agree to the statements and want to participate in the study please click on the '**submit and continue**' button below.

---

**Submit and continue >**

### D.3 Example screenshot from the first round questionnaire

## Delphi Study - Therapist Rating Scale Development (Round 1, Version 1)

41% complete

### Page 6: Rating items

#### This section focuses on interventions to increase awareness

For each item, please rate how much you agree that:

1. The item is **important** and should be included in the scale,
2. The item is **comprehensible**.

Rating are made on a scale of 1-7, where 1 is 'strongly disagree' and 7 is 'strongly agree'.

If you have any comments or suggestions for rewording, please use the boxes provided.

This part of the survey uses a table of questions, [view as separate questions instead?](#)

Consider the item: "**To what extent did the therapist encourage the client to contrast their addictive behaviour with personal goals or values?**" Using the response options in the table below, please indicate the extent you agree that: i) the item is important and should be included in the scale, and ii) the item is comprehensible. If you have any comments or suggestions for altering the item, please detail them in the box provided.

	(1) strongly disagree	(2) disagree	(3) disagree somewhat	(4) neither disagree or agree	(5) agree somewhat	(6) agree	(7) strongly agree
The item is important	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The item is comprehensible	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Comments or suggested rewording:

This part of the survey uses a table of questions, [view as separate questions instead?](#)

Consider the item: **"To what extent did the therapist explore the positive and negative aspects of the client's substance use?"**

	(1) strongly disagree	(2) disagree	(3) disagree somewhat	(4) neither disagree or agree	(5) agree somewhat	(6) agree	(7) strongly agree
The item is important	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The item is comprehensible	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Comments or suggested rewording:

This part of the survey uses a table of questions, [view as separate questions instead?](#)

Consider the item: **"To what extent did the therapist cover both the pros and cons for change?"**

	(1) strongly disagree	(2) disagree	(3) disagree somewhat	(4) neither disagree or agree	(5) agree somewhat	(6) agree	(7) strongly agree
The item is important	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The item is comprehensible	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Comments or suggested rewording:



**D.4 Participants grouped by demographic information to gauge individual participation across the rounds.**

No.	Round	Gender	Residence	Years of experience		Role	Comment
				Therapies	Addictions		
<b>1</b>	1	<b>Male</b>	North America	27	27	Both	Error: R1 gender.
<b>1</b>	2	Female	North America	25	25	Both	
<b>1</b>	3	Female	North America	25	25	Both	
<b>2</b>	1	Male	North America	15	<b>28</b>	Academic	Error: R1 years of addictions experience.
<b>2</b>	2	Male	North America	15	15	Academic	
<b>2</b>	3	Male	North America	16	15	Academic	
<b>3</b>	1	Female	Europe	36	29	Clinician	
<b>3</b>	2	Female	Europe	34	30	Clinician	
<b>3</b>	3	Female	Europe	39	31	Clinician	
<b>4</b>	1	Male	Europe	40	30	Both	
<b>4</b>	2	Male	Europe	35	30	Both	
<b>4</b>	3	Male	Europe	50	40	Both	
<b>5</b>	1	Female	North America	31	31	Both	
<b>5</b>	2	Female	North America	33	32	Both	
<b>5</b>	3	Female	North America	35	35	Both	
<b>6</b>	1	Male	Europe	40	32	Both	R1 only
<b>7</b>	1	Male	Europe	10	33	Academic	R1 only
<b>8</b>	1	Male	Europe	28	34	Both	R1 only
<b>9</b>	1	Male	Europe	40	35	Both	
<b>9</b>	2	Male	Europe	44	42	Both	
<b>9</b>	3	Male	Europe	40	40	Both	
<b>10</b>	1	Female	Europe	17	<b>36</b>	Clinician	Error: R1 years of addictions experience.
<b>10</b>	2	Female	Europe	17	4	Clinician	
<b>10</b>	3	Female	Europe	17	4	Clinician	
<b>11</b>	1	Male	Europe	30	37	Both	R1, R2
<b>11</b>	2	Male	Europe	30	27	Both	
<b>12</b>	1	Female	Europe	20	38	Both	
<b>12</b>	2	Female	Europe	20	25	Both	
<b>12</b>	3	Female	Europe	20	28	Both	
<b>13</b>	2	Female	Europe	0	23	Clinician	R2, R3
<b>13</b>	3	Female	Europe	0	24	Clinician	
<b>14</b>	2	Female	Europe	20	20	Clinician	R2, R3
<b>14</b>	3	Female	Europe	20	20	Clinician	
<b>15</b>	2	Male	Europe	30	0	Both	R2 only

R1 = Round 1; R2 = Round 2; R3 = Round 3.

Error = Potential selection error made by participants when completing the demographic questionnaire. Potential errors were judged by the supervision team using the grouped demographic information, and the number of participants who completed each round of questionnaires: round one ( $n=12$ ); round two ( $n=12$ ); round three ( $n=10$ ).

## D.5 Statistical summary of participants' ratings in round one

Item Reference	Importance Ratings			Comprehensibility Ratings			General	
	% Agreement (% Scoring 6-7)	Median (IQR)	Range	% Agreement (% Scoring 6-7)	Median (IQR)	Range	Number of comments	Outcome
1. Problem focused	92% (83%)	6 (6,7)	3-7	92% (67%)	6 (5,6.75)	2-7	5	Revised
2. Exploring behaviours	75% (42%)	5 (4.25,6.75)	3-7	50% (25%)	4.5 (3,5.75)	2-7	8	<b>Removed<sup>1</sup></b>
3. Exploring substance use	83% (67%)	6 (5,7)	4-7	75% (58%)	6 (4.25,7)	3-7	7	Revised
4. Exploring impact of substance use	100% (92%)	6 (6,7)	5-7	92% (83%)	6 (6,7)	3-7	5	Revised
5. Exploring pros and cons of change	92% (75%)	6 (5.25,7)	4-7	83% (33%)	5 (5,6)	2-7	8	Revised
6. Developing discrepancy	100% (92%)	6 (6,7)	5-7	92% (50%)	5.5 (5,6)	3-7	5	Revised
7. Talking about change	92% (75%)	7 (5.25,7)	3-7	83% (75%)	6 (5.25,7)	3-7	6	<b>Removed<sup>2</sup></b>
8. Treatment goals	100% (92%)	6.5 (6,7)	5-7	83% (75%)	7 (5.25, 7)	2-7	3	Revised
9. Behaviour change planning	92% (83%)	7 (6,7)	3-7	92% (58%)	6 (5,7)	3-7	4	Revised
10. Identifying sources of support	92% (75%)	6 (5.25,7)	2-7	83% (50%)	6 (5.25,7)	2-7	8	Revised
11. Involving others	83% (58%)	6 (5,7)	3-7	67% (58%)	6 (4,6)	3-7	6	<b>Removed<sup>3</sup></b>
12. Homework assigned	83% (67%)	6 (5,7)	2-7	33% (33%)	3.5 (2,6)	1-7	10	Revised
13. Homework reviewed	100% (83%)	7 (6,7)	5-7	92% (92%)	7 (6,7)	4-7	4	Revised
14. Strengths and affirmations	92% (92%)	7 (6,7)	4-7	83% (58%)	6 (5,7)	3-7	4	Revised
15. Collaboration	92% (58%)	6 (5,7)	2-7	50% (33%)	3.5 (2.25,6.75)	2-7	8	Revised
16. Empathy	100% (100%)	7 (6.25,7)	6-7	83% (83%)	6.5 (6,7)	2-7	4	Revised
17. Simple reflections	67% (33%)	5 (3.25,6)	2-6	75% (67%)	6 (4.25,6.75)	2-7	5	<b>Removed<sup>4</sup></b>
18. Complex reflections	92% (75%)	6 (5.25,7)	2-7	75% (33%)	5 (4.25,6)	2-6	7	Revised

% Agreement = percentage of participants scoring between 5 and 7; % Scoring 6-7 = percentage of participants scoring between 6 and 7; IQR = interquartile range, presented as the 25<sup>th</sup>, 75<sup>th</sup> percentiles.

<sup>1</sup> Item 2 'exploring behaviours' removed: Covered by item 4 'exploring impact of substance use', and item 5 'exploring pros and cons of change'.

<sup>2</sup> Item 7 'talking about change' removed: Covered by item 5 'exploring pros and cons of change'.

<sup>3</sup> Item 11 'involving others' removed: Covered by item 10 'Identifying sources of support'.

<sup>4</sup> Item 17 'simple reflections' removed: Covered by item 16 'empathy'.

## D.6 Statistical summary of participants' ratings in round two

Item Reference	Importance Ratings			Comprehensibility Ratings			General	
	% Agreement (% Scoring 6-7)	Median (IQR)	Range	% Agreement (% Scoring 6-7)	Median (IQR)	Range	Number of comments	Outcome
1. Problem focused	92% (67%)	6 (5,7)	4-7	92% (92%)	7 (6,7)	4-7	0	Revised
2. Exploring substance use	75% (50%)	5.5 (4.25,6)	4-7	92% (83%)	7 (6,7)	3-7	3	<b>Removed<sup>1</sup></b>
3. Exploring impact of substance use	83% (58%)	6 (5,7)	3-7	83% (75%)	7 (5.25,7)	4-7	3	<b>Removed<sup>2</sup></b>
4. Exploring pros and cons of change	92% (75%)	7 (5.25,7)	3-7	83% (75%)	7 (5.25,7)	3-7	2	Revised
5. Developing discrepancy	100% (83%)	6 (6,7)	5-7	100% (83%)	6.5 (6,7)	5-7	0	Retained
6. Treatment goals	92% (75%)	7 (5.25,7)	4-7	83% (83%)	7 (6,7)	3-7	4	Revised
7. Behaviour change planning	83% (75%)	6 (5.25,7)	4-7	92% (92%)	7 (6,7)	4-7	3	Revised
8. Identifying sources of support	83 (67%)	6 (5,7)	4-7	83% (67%)	6 (5,7)	4-7	2	Revised
9. Homework assigned	83% (83%)	7 (6,7)	4-7	100% (92%)	7 (6.25,7)	5-7	2	Retained
10. Homework reviewed	92% (83%)	6 (6,7)	4-7	100% (92%)	7 (6,7)	5-7	2	Revised
11. Strengths and affirmations	100% (83%)	7 (6,7)	5-7	92% (75%)	6.5 (5.25,7)	4-7	3	Revised
12. Collaboration	100% (100%)	7 (7,7)	6-7	83% (75%)	6.5 (5.25,7)	3-7	6	Retained
13. Empathy	100% (100%)	7 (7,7)	6-7	100% (100%)	7 (6.25,7)	6-7	1	Retained
14. Complex reflections	100% (75%)	6 (5.25,7)	5-7	67% (58%)	6 (3.25,7)	1-7	5	Revised

% Agreement = percentage of participants scoring between 5 and 7; % Scoring 6-7 = percentage of participants scoring between 6 and 7; IQR = interquartile range, presented as the 25<sup>th</sup>, 75<sup>th</sup> percentiles.

<sup>1</sup> Item 2 'exploring substance use' removed: Overlap with item 4 'exploring pros and cons of change' and item 14 'complex reflections'.

<sup>2</sup> Item 3 'exploring impact of substance use' removed: Overlap with item 4 'exploring pros and cons of change'.

## D.7 The BATS



### Brief Addiction Therapist Scale

A tool for evaluating therapists' delivery of psychological therapies for alcohol and drug use problems.

Designed to facilitate training and supervision, and enhance therapist skill.

**Instructions for use:** For each item, circle a number on the 5-point scale reflecting the extent to which the therapist carried out the behaviour. For items that are not applicable to the session, score 0 'not at all'. Use the space provided on page 2 to give context, comments, and additional information e.g. the client's first session. Item definitions are provided on page 2. To be used with audio or video recordings of therapy sessions.

During the session...	Not at all	A little	Somewhat	A good deal	Extensively
1. The therapist kept the session focused on the aims for that session.	0	1	2	3	4
2. The therapist attempted to work together with the client.	0	1	2	3	4
3. The therapist conveyed empathy.	0	1	2	3	4
4. The therapist focused on the client's strengths.	0	1	2	3	4
5. The therapist used "complex reflections" – offering a perspective which added meaning and enabled the client to make connections.	0	1	2	3	4
6. The therapist and the client planned tasks for the client to do between sessions.	0	1	2	3	4
7. The therapist and the client reviewed tasks planned in the previous session.	0	1	2	3	4
8. The therapist enabled the client's goals for treatment to be discussed.	0	1	2	3	4
If in this session the focus was on building motivation for change:					
9. The therapist encouraged the client to consider inconsistencies between their substance use, and personal goals or values.	0	1	2	3	4
10. The therapist encouraged the client to talk about the positive aspects of changing substance use.	0	1	2	3	4
If in this session the focus was on planning or maintaining change:					
11. The therapist enabled a plan for changing the client's substance use, or maintaining change, to be discussed.	0	1	2	3	4
12. The therapist discussed how the client's social network might support changing substance use or maintaining change.	0	1	2	3	4

Total score:	
--------------	--

---

### Item definitions

---

1. **Session aims:** The therapist kept the session focused on clinically relevant aims during the session, e.g. target behaviour. This may or may not include explicit discussion of the purpose of the session, e.g. *to describe a relapse prevention plan*. Aims may change during the course of the session following disclosure of risk.
2. **Working together:** Developing a collaborative relationship between the client and the therapist. It is about discussing, actively seeking the client's input; not telling, and not arguing.
3. **Convey empathy:** Making efforts to convey warmth and understanding of the client's thoughts and feelings. The therapist avoids any blaming or labelling.
4. **The client's strengths:** Helping the client to identify and focus on what they can do, not what they cannot do: achievements rather than failings.
5. **Complex reflections:** Helping the client to gain insight by making and/or strengthening connections between things they have said. Going beyond repeating or slightly rephrasing what the client has said.
6. **Planning tasks:** Any task that is planned (the therapist and the client agreed what to do and how to do it) for the client to do between sessions, e.g. *specific homework tasks, trying new behaviours*.
7. **Reviewing tasks:** Explicit discussion in which tasks set in the previous session are reviewed. This item is not applicable if it is the client's first session, tick the box as appropriate.
8. **Treatment goals:** Goals refer to the overall treatment goals, e.g. *abstinence, harm reduction, moderation*. The goals could be discussed by the therapist and/or the client.
9. **Considering inconsistencies:** Exploring how the client's behaviour conflicts with his/her personal goals and values, e.g. *I need to drink a bottle of gin but I want to be a good parent*.
10. **Talking about change:** The therapist encourages the client to talk about the positive aspects of changing.
11. **Change planning:** Discussion of an overall plan to achieve the agreed treatment goals. Tasks represent the steps in the plan to achieve the overall treatment goals.
12. **The social network:** The therapist facilitates a discussion about the client's actual and/or potential social network to identify how this may support the overall plan.

---

### Context, comments, and additional information:

---

©University of Leeds 2017.

We acknowledge the support of Leeds and York Partnership NHS Foundation Trust (LYPFT). BATS was funded by Alcohol Research UK (ARUK), and the Society for the Study of Addiction (SSA). We thank the participants involved in developing BATS.



## **Appendix E**

### **Chapter 6 Appendices**

Appendix D focuses on the appendices for chapter 5, agreeing the content. The following appendices are included:

- i) Therapist information sheet for collecting routine practice recordings (p.261).
- ii) Therapist consent form for collecting routine practice recordings (p.265).
- iii) Therapist questionnaire for collecting routine practice recordings (p.266).
- iv) Client covering letter for collecting routine practice recordings (p.267).
- v) Client information sheet for collecting routine practice recordings (p.268).
- vi) Client consent form for collecting routine practice recordings (p.271).
- vii) Friends and family information sheet for collecting routine practice recordings (p.272).
- viii) Client questionnaire for collecting routine practice recordings (p.275).
- ix) Ordinary least squares and weighted least squares methods (p.276).
- x) Matching Working Alliance Inventory (WAI) data with corresponding sessions in the BATS dataset (p.278).
- xi) Calculating Kendall's tau-b for the BATS and the WAI-S client (p.279).
- xii) Selected items from the trial process rating measures for the convergent validity analyses (p.281).
- xiii) Checks of normality for data used to examine convergent validity (p.284).
- xiv) Calculating weighted kappa for item 1 'problem focused' (p.285).
- xv) Calculating intraclass correlation coefficient for item 1 'problem focused' (p.288).
- xvi) Sample size estimation for the intraclass correlation coefficient (p.291).

## E.1 Routine practice recordings therapist information sheet

### Participant Information Sheet: An Overview

*We are inviting you to take part in our research study.  
Before you decide whether to take part please read this information sheet.  
The Researcher will go through this sheet with you and answer any questions you have.*

#### What is the study about?

- The purpose of the study is to develop the Addiction Therapist Rating Scale (ATRS). This is a tool that can be used by therapists and supervisors in every day practice to measure how well they deliver effective treatments for alcohol and drug use problems.

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

- You will be asked to complete a consent form.
- You will be asked to complete a short questionnaire asking about your experience in the addiction field, level of education and relevant training.
- You will be asked to record your treatment sessions. This will involve identifying potential participants, seeking written consent, recording your treatment sessions onto DVD, and completing a short questionnaire about service users' primary problem substance and treatment history.

#### Do I have to take part?

- No, it is up to you whether or not you decide to take part.

#### Can I change my mind?

- Yes, you are free to withdraw at any time up until data analysis begins, without giving a reason, this includes after you have completed the consent form.

#### Will my taking part in this study be kept confidential?

- Yes, any information you provide will be treated in confidence. You will not be identified in any reports, publications or presentations.

#### Who can I contact?

If you would like further information or have questions about the study please contact:

Helen Crosby  
Researcher  
Email: [umhfc@leeds.ac.uk](mailto:umhfc@leeds.ac.uk)

or

Gillian Tober  
Project Supervisor  
Email: [gillian.tober@nhs.net](mailto:gillian.tober@nhs.net)

\* The Addiction Therapist Scale (ATRS) was renamed the Brief Addiction Therapist Scale (BATS) based on feedback given at the Leeds Institute of Health Sciences Postgraduate Research Symposium (2016).

## Participant Information Sheet: Additional Information

### What is the study about?

The purpose of the study is to develop the Addiction Therapist Rating Scale (ATRS). This is a tool that can be used by therapists and supervisors in every day practice to measure how well they deliver effective treatments for alcohol and drug use problems. The ATRS is being developed to support training and supervision; this has the potential to improve practice and impact on treatment outcomes. We are using video recordings of treatment sessions to test its reliability (the extent to which it produces consistent results) and validity (how well it measures what it purports to measure). We will then implement the ATRS to look at how it is used in addiction services.

### Why have I been approached?

You have been approached because you work at the [*specialist addictions service (SAS)*], where treatment sessions are routinely recorded for supervision purposes. In this study, we would like to collect and use these recordings to test whether the ATRS is a reliable and valid measure for rating the delivery of effective treatments for alcohol and drug use problems.

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

If you agree to take part, you will be asked to complete a consent form and a short questionnaire asking about your experience in the addiction field, level of education and relevant training. You will then be invited to record your treatment sessions. This will involve:

- (i) **Identifying service users.** You will be asked to liaise with your team secretary to identify service users on your current caseload who can be approached to take part in the study. Your team secretary will post an information pack on your behalf to service users who you identify as eligible for inclusion and who will be attending therapy sessions the following week. The information pack will include a copy of the service user participant information sheet and a covering letter introducing the study and your involvement.
- (ii) **Explaining the study.** You will be given copies the service user participant information sheet and consent form. At the start of each treatment session, go through the service user participant information sheet highlighting what the study is about and what it will involve.
- (iii) **Taking consent.** If service users agree to their sessions being recorded (without appearing in the shot), they will need to complete a service user consent form. A copy of the completed form should be offered for their own records. If service users do not consent, no further involvement is required.
- (iv) **Recording the session.** If service users consent, record the treatment session focusing the recording device on the therapist; this means that the service user will be heard in audio only. After the session, finalise the recording onto DVD as described in the [*SAS*] Audio Video instructions.



- (v) **Completing the service user questionnaire.** You will be given copies of the service user questionnaire. The questionnaire asks about service users' primary problem substance and treatment history, data routinely collected at the [SAS]. Complete the questionnaire for each service user.
- (vi) **Returning the paperwork.** Return the DVD and paperwork to Gillian ~~Tober's~~ office at the [SAS].

**Do I have to take part?**

No, it is up to you whether or not you decide to take part. We will describe the study and go through this information sheet. If you decide to take part, we will ask you to sign a consent form. You are free to withdraw at any time up until data analysis begins, this includes after you have completed the consent form. If you chose to withdraw your consent, you do not have to give a reason and your employment status or legal rights will not be affected. If you would like more information please feel free to contact us, our details are at the end of this information sheet. If you do want to take part, please keep this information sheet for your reference.

**Who is being asked to take part?**

Therapists and service users at the [*specialist addictions service*].

**Will my taking part in this study be kept confidential?**

Any information you provide will be treated in confidence. At the beginning of the study we will keep a copy of your signed consent form. The questionnaire will not include your name or any other personally identifiable information; instead we will mark them with a code number. Recordings will be identified by the same code number. Completed consent forms will be stored separately from the questionnaires and video recordings. All information collected during the study will be stored securely, only accessible to authorised members of staff. The recordings will be used throughout the project and in future research. Research findings may include anonymised quotes from those taking part. If you consent, regulatory authorities such as the NHS Trust where the study is being conducted may look at the information collected to check that the study is being carried out correctly. All data will be archived for a period of 6 years following the end of the study and then disposed of securely.

**What will happen to the results of the study?**

The recordings that you provide will be rated by the research team to test whether the ATRS is a reliable and valid measure for rating the delivery of effective treatments for alcohol and drug use problems. The results will be presented as a report and published in health journals. We may also present the findings at conferences and other relevant meetings. However, you will not be identified in any reports, publications or presentations. If you would like to receive a copy of the findings, please contact the research team.

**What are the possible risks of taking part?**

Taking part in this study will involve some of your time. There are no other risks to taking part.

**Who is organising and funding the study?**

The research team comprises: (i) Helen Crosby, Researcher at the [*specialist addictions service*], (ii) Gillian Tober, Consultant Addiction Psychologist at the [*specialist addictions service*], (iii) Gary Latchford, Research Director at the Leeds Institute of Health Sciences (LIHS), University of Leeds, and (iv) Bridgette Bewick, Associate Professor at the LIHS, University of Leeds. The study is being carried out as part of a PhD research project at the University of Leeds and is funded by Alcohol Research UK and the Society for the Study of Addiction.

**What if there is a problem?**

If you have any concerns, or if you require additional support and/or advice in relation to taking part in this study, please contact the Chief Investigator:

Helen Crosby  
Leeds Institute of Health Sciences  
University of Leeds  
Room G.02, Charles ~~Thackrah~~ Building  
101 Clarendon Road  
Leeds  
LS2 9NG  
Email: [umhfc@leeds.ac.uk](mailto:umhfc@leeds.ac.uk)

**Who has reviewed the study?**

This research has been reviewed and approved by Yorkshire and the Humber Leeds West Research Ethics Committee.

**Useful numbers:**

Alcoholics Anonymous	0845 796 7555
Addiction Dependency Solutions	0113 247 0111
St. Anne's Alcohol and Drug Services	0113 243 4486

**Thank you for taking time to read this information sheet.**

## E.2 Routine practice recordings therapist consent form



Participant ID

### Development of the Addiction Therapist Rating Scale (ATRS)

#### Treatment Session Recordings – Therapist Consent Form

##### The ATRS project

Please initial each box

I confirm that I have read and understood the information sheet dated 25/02/2015 (Version 2) for the above study.

I confirm that I have had the opportunity to ask questions about the study and any questions I have asked have been answered satisfactorily.

I understand that my participation is voluntary and that I am free to withdraw at any time up until analysis begins without giving a reason, without my employment status or legal rights being affected.

I confirm that I am willing to allow my treatment sessions to be recorded onto DVD for research purposes.

I agree to the information I disclose as part of the study being used anonymously in the presentation of research findings.

I understand that relevant sections of my data collected during the study may be looked at by individuals from regulatory authorities. I give permission for these individuals to have access to this data.

I agree to take part in the above study.

\_\_\_\_\_  
Name of participant (please print)      Date (dd/mm/yyyy)      Signature of participant

\_\_\_\_\_  
Name of Researcher (please print)      Date (dd/mm/yyyy)      Signature of Researcher

### E.3 Routine practice recordings therapist questionnaire



Participant ID (research use only):

#### Development of the Addiction Therapist Rating Scale (ATRS)

##### Staff Questionnaire

1. How long have you worked in the addictions field (please specify)?

Years       Months (if known)

2. What is your job role (please tick)?

- Psychiatrist
- Psychologist
- Nurse
- Therapist
- Health Support Worker
- Other (please specify): \_\_\_\_\_

3. How long have you worked in your current role (please specify)?

Years       Months (if known)

4. What qualifications do you have (please specify)?

---

---

---

5. What relevant training have you completed (please specify)?

---

---

---

## E.4 Routine practice recordings client covering letter

Leeds and York Partnership   
NHS Foundation Trust

**[Specialist addictions service (SAS)]**

[SAS address]

Telephone: [SAS telephone number]

[Date]

[Service user name]

[Service user address]

Dear [Name],

**Re: Information about research being conducted at [SAS]**

I am one of the therapists involved in a study investigating how therapists might measure how well they deliver treatments for alcohol and drug use problems.

I am writing to you with information about the study so that you can have opportunity to read the information before our appointment next week. Whether or not you decide to participate in the study will not affect your treatment.

I look forward to seeing you next week.

Yours sincerely,

[Therapist name]

Study 2a Cover Letter 23-02-2013 [Version 1]

## E.5 Routine practice recordings client information sheet

### Participant Information Sheet: An Overview

*We are inviting you to take part in our research study.  
Before you decide whether to take part please read this information sheet.  
Your keyworker will go through this sheet with you.*

#### What is the study about?

- We would like to record your appointment to help us to develop a scale that can be used by therapists in every day practice to measure how well they deliver treatments for alcohol and drug use problems.

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

- You will be asked to complete a consent form.
- Your keyworker will record your appointment.
- Your keyworker will provide us with information on your age, gender, **primary** problem substance and treatment history for your addiction problem.

#### Will I be seen in the recording?

- No, the recorder will focus on your keyworker.
- Your voice will be heard but you will not be seen in the recording.

#### Do I have to take part?

- No, it is up to you whether or not you decide to take part.

#### Can I change my mind?

- Yes, you can withdraw your consent up until analysis begins, without giving a reason and this will not affect the care you receive. Once we have analysed the recordings we will not be able to remove data that is associated with your recording from the study.

#### Will my taking part in this study be kept confidential?

- Yes, any information you provide will be treated in confidence. You will not be identified in any reports, publications or presentations.

#### Who can I contact?

If you would like further information or have questions about the study please contact:

Helen Crosby  
Researcher

or

Gillian Tober  
Project Supervisor

Email: [helen.crosby@nhs.net](mailto:helen.crosby@nhs.net)

Email: [gillian.tober@nhs.net](mailto:gillian.tober@nhs.net)

[Specialist addictions service address and telephone number]

## Participant Information Sheet: Additional Information

### What is the study about?

The purpose of the study is to develop the Addiction Therapist Rating Scale (ATRS). This is tool that can be used by therapists in every day practice to measure how well they deliver treatments for alcohol and drug use problems. The scale is being designed to support training and routine supervision, which has the potential to improve practice and impact on patient outcomes. We are using video recordings of treatment sessions to test its reliability (the extent to which it produces consistent results) and validity (how well it measures what it purports to measure). We will then implement the scale to look at how it is used in addiction services.

### Why have I been approached?

You have been approached because you have attended an appointment with your keyworker. We would like to record your treatment to help us look at whether the scale is a good measure for rating how well therapists deliver treatments for alcohol and drug use problems.

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

If you agree to take part, you will be asked to complete a consent form and your keyworker will record your appointment. Because we are interested in seeing what therapists say and do, the recorder will only focus on your keyworker. This means that you will **not** appear on the DVD. Instead, only your voice will be heard. You can ask your keyworker to stop the recording at any point. When we write up the results, it is helpful to describe who took part in the study. If you agree, we will ask you therapists to provide us with information on your age, gender, primary problem substance and treatment history for your addiction problem. This information will be taken from your medical notes.

### Do I have to take part?

No, it is up to you whether or not you decide to take part. Your keyworker will describe the study and explain this information sheet. If you decide to take part, we will ask you to sign a consent form. You are free to withdraw your consent up until analysis begins, without giving a reason, and this will not affect the care you receive. If you would like more information please contact us, our details are at the end of this information sheet. If you do want to take part, please keep this information sheet for your reference.

### Will my taking part in this study be kept confidential?

Any information you provide will be treated in confidence. At the beginning of the study we will keep a copy of your signed consent form. Recordings and questionnaire data will be identified by a code number only. Completed consent forms will be stored separately from the participant information and video recordings. All information will be stored securely, only accessible to authorised members of staff. The recordings will be used throughout the project and in future research. Research findings may include anonymised quotes from those taking part. If you consent, regulatory authorities such as the NHS Trust where the study is being conducted may

look at the information collected to check that the study is being carried out correctly. All data will be archived for a period of 6 years following the end of the study and then disposed of securely.

**What are the possible risks of taking part?**

Taking part in this study will involve some of your time. There are no other risks to taking part. If you do not wish to take part, your treatment and future care will not be affected.

**Who is being asked to take part?**

Therapists and service users at the [*specialist addictions service*].

**What will happen to the results of the study?**

The recordings will be used by the research team to see whether the scale is a good measure for rating how well therapists follow correct treatment procedures. Information on your primary problem substance and treatment history will be used to describe who took part in the study. The results will be presented as a report and published in health journals. We may also present the findings at conferences and other relevant meetings. You will not be identified in any reports, publications or presentations. If you would like to receive a copy of the findings please contact the research team.

**What if there is a problem?**

If you have a concern about any aspect of this study, please speak to the research team. If you remain unhappy, you can go to the Patient Advice and Liaison Service (PALS), who can be contacted on 0800 0525 790 or at [pals.lypft@nhs.net](mailto:pals.lypft@nhs.net) for further guidance.

**Who is organising and funding the study?**

The research team comprises: (i) Helen Crosby, Researcher at the [*specialist addictions service*], (ii) Gillian Tober, Consultant Addiction Psychologist at the [*specialist addictions service*], (iii) Gary Latchford, Research Director at the Leeds Institute of Health Sciences (LIHS), University of Leeds, and (iv) Bridgette Bewick, Associate Professor at the LIHS, University of Leeds. The study is being carried out as part of a PhD research project at the University of Leeds and is funded by Alcohol Research UK and the Society for the Study of Addiction.

**Who has reviewed the study?**

This research has been reviewed and approved by Yorkshire and the Humber Leeds West Research Ethics Committee.

**Useful numbers:**

Alcoholics Anonymous	0845 796 7555
Addiction Dependency Solutions	0113 247 0111
St. Anne's Alcohol and Drug Services	0113 243 4486

**Thank you for taking time to read this information sheet.**

The ATRS project: Study 2 Treatment Session Recordings Service User PIS 22/09/2014 (Version 1)



## E.6 Routine practice recordings client consent form



Participant ID

### Development of the Addiction Therapist Rating Scale (ATRS)

#### Treatment Session Recordings – Service User Consent Form

##### The ATRS project

Please initial each box

I confirm that I have read and understood the information sheet dated 22/09/2014 (Version 1) for the above study.

I confirm that I have had the opportunity to ask questions about the study and any questions I have asked have been answered satisfactorily.

I understand that my participation is voluntary and that I am free to withdraw at any time up until data analysis begins without giving a reason, without my medical care or legal rights being affected.

I confirm that I am willing to allow my treatment sessions to be recorded onto DVD (without me appearing in the shot) for research purposes.

I confirm that I am willing to allow my keyworker to provide information on my primary problem substance and treatment history for research purposes.

I understand that relevant sections of my data collected during the study may be looked at by individuals from the research team or from regulatory authorities. I give permission for these individuals to have access to this data.

I agree to take part in the above study.

\_\_\_\_\_  
Name of participant (please print)

\_\_\_\_\_  
Date (dd/mm/yyyy)

\_\_\_\_\_  
Signature of participant

\_\_\_\_\_  
Name of keyworker (please print)

\_\_\_\_\_  
Date (dd/mm/yyyy)

\_\_\_\_\_  
Signature of keyworker

## E.7 Routine practice recordings friends and family information sheet

### Friends and Family Information Sheet: An Overview

*We are inviting you to take part in our research study.  
Before you decide whether to take part please read this information sheet.  
The keyworker will go through this sheet with you.*

#### What is the study about?

- We would like to record the treatment session you have attended to help us to develop a scale that can be used by therapists in every day practice to measure how well they deliver treatments for alcohol and drug use problems.

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

- You will be asked to complete a consent form.
- The keyworker will record your appointment.

#### Will I be seen in the recording?

- No, the recorder will focus on the keyworker.
- Your voice will be heard but you will not be seen in the recording.

#### Do I have to take part?

- No, it is up to you whether or not you decide to take part.


#### Can I change my mind?

- Yes, you can withdraw your consent up until analysis begins, without giving a reason. Once we have analysed the recordings we will not be able to remove data that is associated with your recording from the study.

#### Will my taking part in this study be kept confidential?

- Yes, any information you provide will be treated in confidence. You will not be identified in any reports, publications or presentations.

#### Who can I contact?

 If you would like further information or have questions about the study please contact:

Helen Crosby

or

Gillian Tober

Researcher

Project Supervisor

Email: [helen.crosby@nhs.net](mailto:helen.crosby@nhs.net)

Email: [gillian.tober@nhs.net](mailto:gillian.tober@nhs.net)

[Specialist addictions service address and telephone number]

## Participant Information Sheet: Additional Information

### What is the study about?

The purpose of the study is to develop the Addiction Therapist Rating Scale (ATRS). This is tool that can be used by therapists in every day practice to measure how well they deliver treatments for alcohol and drug use problems. The scale is being designed to support training and routine supervision, which has the potential to improve practice and impact on patient outcomes. We are using video recordings of treatment sessions to test its reliability (the extent to which it produces consistent results) and validity (how well it measures what it purports to measure). We will then implement the scale to look at how it is used in addiction services.

### Why have I been approached?

You have been approached because you have attended an appointment with your friend or family member. We would like to record the treatment session to help us look at whether the scale is a good measure for rating how well therapists deliver treatments for alcohol and drug use problems.

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

If you agree to take part, you will be asked to complete a consent form. If your friend or family member also agrees to take part, the keyworker will record your appointment. Because we are interested in seeing what therapists say and do, the recorder will only focus on the keyworker. This means that you will **not** appear on the DVD. Instead, only your voice will be heard. You can ask the keyworker to stop the recording at any point.

### Do I have to take part?

No, it is up to you whether or not you decide to take part. The keyworker will describe the study and explain this information sheet. If you decide to take part, we will ask you to sign a consent form. You are free to withdraw your consent up until analysis begins, without giving a reason. If you would like more information please contact us, our details are at the end of this information sheet. If you do want to take part, please keep this information sheet for your reference.

### Will my taking part in this study be kept confidential?

Any information you provide will be treated in confidence. At the beginning of the study we will keep a copy of your signed consent form. Recordings will be identified by a code number only. Completed consent forms will be stored separately from the video recordings. All information will be stored securely, only accessible to authorised members of staff. The recordings will be used throughout the project and in future research. Research findings may include anonymised quotes from those taking part. If you consent, regulatory authorities such as the NHS Trust where the study is being conducted may look at the information collected to check that the study is being carried out correctly. All data will be archived for a period of 6 years following the end of the study and then disposed of securely.

**What are the possible risks of taking part?**

Taking part in this study will involve some of your time. There are no other risks to taking part.

**Who is being asked to take part?**

Therapists, service users, and friends and/or family members who attend an appointment at the [specialist addictions service].

**What will happen to the results of the study?**

The recordings will be used by the research team to see whether the scale is a good measure for rating how well therapists follow correct treatment procedures. The results will be presented as a report and published in health journals. We may also present the findings at conferences and other relevant meetings. You will not be identified in any reports, publications or presentations. If you would like to receive a copy of the findings please contact the research team.

**What if there is a problem?**

If you have a concern about any aspect of this study, please speak to the research team. If you remain unhappy, you can go to the Patient Advice and Liaison Service (PALS), who can be contacted on 0800 0525 790 or at [pals.lypft@nhs.net](mailto:pals.lypft@nhs.net) for further guidance.

**Who is organising and funding the study?**

The research team comprises: (i) Helen Crosby, Researcher at the [specialist addictions service], (ii) Gillian Tober, Consultant Addiction Psychologist at the [specialist addictions service], (iii) Gary Latchford, Research Director at the Leeds Institute of Health Sciences (LIHS), University of Leeds, and (iv) Bridgette Bewick, Associate Professor at the LIHS, University of Leeds. The study is being carried out as part of a PhD research project at the University of Leeds and is funded by Alcohol Research UK and the Society for the Study of Addiction.

**Who has reviewed the study?**

This research has been reviewed and approved by Yorkshire and the Humber Leeds West Research Ethics Committee.

**Useful numbers:**

Alcoholics Anonymous	0845 796 7555
Addiction Dependency Solutions	0113 247 0111
St. Anne's Alcohol and Drug Services	0113 243 4486

**Thank you for taking time to read this information sheet.**

## E.8 Routine practice recordings client questionnaire

Participant ID

### Development of the Addiction Therapist Rating Scale (ATRS)

#### Service User Questionnaire

#### The ATRS project



1. What is the service user's age?

Years

2. Is the service user male or female? (please tick as appropriate)

Male  Female

3. What do you see as the most problematic drug(s), including alcohol and recreational drugs, for the service user?

	Substance	Age 1st used	Days in the week 0 -7	Reasons/events (if any) leading to use
1				
2				
3				

4. Has the service user had treatment for their addiction problem before?

If yes, what treatment have they had? .....

Yes  No

## E.9 Ordinary least squares and weighted least squares methods

In regression, the relationship between two variables is defined by as:

$$Y = b_o + b_iX + e$$

Where:

$Y$  = The dependent or outcome variable.

$X$  = The independent or predictor variable.

$b_o$  = The constant coefficient, the value of  $Y$  when  $X$  is zero (the intercept).

$b_i$  = The regression coefficient for  $X$ , the amount by which  $X$  would increase for every one unit increase of  $Y$  (the slope) (Bowers, 2014).

$e$  = The error or residual term, the amount of variability not explained by the relationship with  $X$  (Bowers et al., 2014; Bland, 2017).

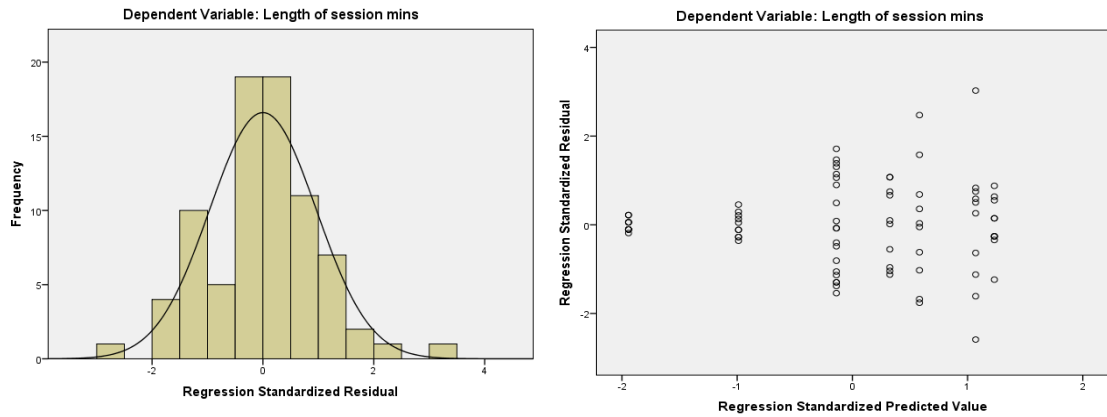
The regression coefficients can be estimated using the method of ordinary least squares (OLS). If the values for  $Y$  and  $X$  are plotted on a scatterplot, a regression line that best fits the data can be drawn (Bowers et al., 2014). Inevitably there will be random variation; the plotted points will deviate from the line. The distance of each point from the regression line represents the error term (Bowers, 2014). OLS finds the best line by minimising “the sum of all the squared error terms (i.e. minimises  $\sum e^2$ )” (Bowers et al., 2014, p.177). The OLS method has two main assumptions: the residuals are normally distributed, and the variance of the residuals are the same for each value of  $X$  (Bland, 2017).

A linear regression model was used to compare the length of sessions for each therapy type. Assumptions for the OLS method were considered; the residuals were normally distributed, but the variance of the residuals was not the same (**Figure 1**). Violating the assumption of homoscedasticity invalidates the standard error of the regression coefficient (Bland, 2017). As such, the confidence intervals (CIs) and t-tests will be biased as they are computed using the standard error (Field, 2017). The analysis was, therefore, rerun using a weighted least squares (WLS) regression model (Field, 2017). WLS applies a transformation to the original OLS model by multiplying both sides of the regression equation by a weight ( $w_t$ ):

$$w_tY = w_tb_o + w_tb_iX + w_te$$

The weights are inversely proportional to the standard deviation of the residuals, and can be represented as  $w_t = \frac{1}{\sigma^2}$  where  $\sigma^2$  is the error variance (NIST/SEMATECH, 2013). The

observations (or plotted points) with a smaller error variance are given more weight than the observations with a larger error variance. The weights have to be estimated, the actual error variance is unknown. Weights can be calculated by regressing the absolute values of the residuals against the independent variable(s); the fitted values from the regression analysis are considered estimates of error variance, and enable the weights to be calculated.



**Figure 1: Length of sessions – checks of normality and variance for the OLS method**

## E.10 Matching WAI data with corresponding sessions in the BATS dataset

Sessions rated using the BATS				WAI client data	WAI therapist data
Trial	Session ID	Therapy	Session number	Session number	Session number
ADAPTA	1001	AF	1	-	-
	1004	HL	2	3 <sup>a</sup>	-
	1004	HL	1	-	3 <sup>a</sup>
	1007	AF	3	3	3
	1009	HL	2	2	2
	1018	AF	2	2	2
	1025	HL	3	-	3
	1036	HL	2	3 <sup>a</sup>	2
	1037	AF	4	3	3
	1040	AF	3	3	3
	1047	AF	4	3	3
	1054	AF	1	2	2
	1058	AF	1	-	-
	1062	AF	1	-	-
	1075	HL	1	2	2
	1080	HL	3	3	2 <sup>b</sup>
	1080	HL	1	-	-
	1082	AF	2	2	2
	1084	HL	2	2	2
1086	HL	3	2 <sup>b</sup>	3	
			(n=20)	(n=14)	(n=15)
UKATT	01-0394-1225	SBNT	6	1 <sup>c</sup>	8
	01-0524-1351	SBNT	7	8	8
	01-0613-1424	MET	1	1	1
	02-0063-1042	SBNT	1	1	1
	02-0087-1210	MET	1	1	1
	02-0421-1778	MET	2	1	1
	03-0030-1141	MET	1	-	-
	04-0034-1216	MET	2	1	1
	04-0039-1249	MET	2	3 <sup>d</sup>	3 <sup>d</sup>
	04-0101-1648	MET	2	1	1
	06-0041-1008	SBNT	5	-	-
	08-0170-1139	MET	3	-	1 <sup>c</sup>
	08-0326-1352	SBNT	1	-	-
	08-0654-1601	SBNT	4	8 <sup>d</sup>	1
	08-0800-1763	SBNT	7	-	1 <sup>c</sup>
	08-0664-1616	MET	3	3	3
	08-0716-1680	MET	3	3	8
	08-0732-1708	SBNT	4	-	1
	08-0764-1729	SBNT	3	1	1
	09-0036-1222	SBNT	3	1	8 <sup>d</sup>
01-0394-1225	SBNT	6	8	8	
			(n=20)	(n=15)	(n=18)

<sup>a</sup> WAI data collected at the end of session 3 was used, missing session 2 data (ADAPTA).

<sup>b</sup> WAI data collected at the end of session 2 used, missing session 3 data (ADAPTA).

<sup>c</sup> WAI data collected at the end of the first session used, missing last session data (UKATT).

<sup>d</sup> WAI data collected at the end of the last session used, missing first session data (UKATT).



### E.11 Calculating Kendall's tau-b for the BATS and the WAI-S client.

WAI data from ADAPTA and UKATT were used to examine convergent validity of the BATS.

Data derived from the BATS and the WAI-S client are presented in **Table 1** ( $n=24$ ).

**Table 1: Scores on the BATS and the WAI-S client**

Case	Trial	BATS Data		WAI-S Client Data	
		Total Scores	Ranked Scores	Total Scores	Ranked Scores
1	UKATT	7	1	62	14
2	UKATT	10	2	55	9
3	UKATT	11	3	54	7
4	UKATT	12	4.5	38	1
5	UKATT	12	4.5	51	3
6	UKATT	13	6	54	7
7	UKATT	14	7	58	10.5
8	UKATT	15	8.5	52	4
9	UKATT	15	8.5	53	5
10	ADAPTA	16	10	78	20
11	UKATT	17	11	58	10.5
12	UKATT	19	12	54	7
13	ADAPTA	21	13	77	17.5
14	ADAPTA	22	14	81	23
15	ADAPTA	23	15.5	61	12.5
16	ADAPTA	23	15.5	72	15
17	ADAPTA	25	17	80	22
18	ADAPTA	29	18	75	16
19	UKATT	30	19.5	61	12.5
20	ADAPTA	30	19.5	78	20
21	UKATT	32	21.5	44	2
22	ADAPTA	32	21.5	78	20
23	ADAPTA	33	23	77	17.5
24	ADAPTA	36	24	83	24

Kendall's tau-b is a measure of association based on the number of concordant and discordant pairs. A concordant pair is when the scores for each case are ordered in the same way by both variables, in this case the BATS and the WAI-S client. A discordant pair is when the scores are ordered in opposite ways (Bland, 2000). Cases that have the same value on one or both variables are called a tied pair (or tied rank). **Table 2** shows the number of concordant and discordant pairs.

Kendall's tau-b is defined by Bland (2000) as:

$$\tau_b = \frac{n_c - n_d}{\sqrt{\left(\frac{n(n-1)}{2} - \sum \frac{t(t-1)}{2}\right) \left(\frac{n(n-1)}{2} - \sum \frac{u(u-1)}{2}\right)}}$$

Where:

- $n_c$  = Number of concordant pairs ( $n=192$ ),
- $n_d$  = Number of discordant pairs ( $n=84$ ),
- $n$  = Number of cases ( $n=24$ ),
- $t$  = Number of cases tied at a particular rank for the variable WAI-S client,
- $u$  = Number of cases tied at a particular rank for the variable BATS.

**Table 2: Number of concordant and discordant pairs using WAI-S client ranked scores**

<b>Rank*:</b>	<b>14</b>	<b>9</b>	<b>7</b>	<b>1</b>	<b>3</b>	<b>7</b>	<b>10.5</b>	<b>4</b>	<b>5</b>	<b>20</b>	<b>10.5</b>	<b>7</b>
$n_c$	10	14	15	20	18	14	12	15	14	4	11	11
$n_d$	13	8	6	0	1	4	5	1	1	10	2	1
<b>Rank*:</b>	<b>17.5</b>	<b>23</b>	<b>12.5</b>	<b>15</b>	<b>22</b>	<b>16</b>	<b>12.5</b>	<b>20</b>	<b>2</b>	<b>20</b>	<b>17.5</b>	<b>24</b>
$n_c$	5	1	7	6	1	4	4	1	3	1	1	0
$n_d$	6	9	2	2	6	2	1	3	0	1	0	0

\* Ranked scores from the WAI-S client data.

$n_c$  = Number of concordant pairs;  $n_d$  = Number of discordant pairs.

The definition shows how Kendall's tau-b adjusts for ties. The groups of tied ranks for each variable contribute to  $n_c - n_d$ , the difference between the number of concordant and discordant pairs (Bland, 2000). **Table 3** shows the tied ranks and their contribution to Kendall's tau-b. The relationship between total scores on the BATS and the WAI-S client is calculated as:

$$\tau_b = \frac{192 - 84}{\sqrt{\left(\frac{24(24-1)}{2} - 9\right)\left(\frac{24(24-1)}{2} - 5\right)}} = \frac{108}{\sqrt{(276-9)(276-5)}} = 0.401$$

**Table 3: Calculating  $\sum t(t-1)/2$  and  $\sum u(u-1)/2$**

Variable	Tied Rank	No. Cases	Equation	Total
WAI-S Client	7	3	$3 \times (3 - 1)/2$	3
	10.5	2	$2 \times (2 - 1)/2$	1
	12.5	2	$2 \times (2 - 1)/2$	1
	17.5	2	$2 \times (2 - 1)/2$	1
	20	3	$3 \times (3 - 1)/2$	3
			<b><math>\sum t(t-1)/2</math></b>	<b>9</b>
BATS	4.5	2	$2 \times (2 - 1)/2$	1
	8.5	2	$2 \times (2 - 1)/2$	1
	15.5	2	$2 \times (2 - 1)/2$	1
	19.5	2	$2 \times (2 - 1)/2$	1
	21.5	2	$2 \times (2 - 1)/2$	1
			<b><math>\sum u(u-1)/2</math></b>	<b>5</b>

No. Cases = Number of cases.

## E.12 Selected items from the trial process rating measures for the convergent validity analyses

BATS Item Reference	ADAPTA PRS <sup>1</sup> Item Reference: Item	AESOPS PRS <sup>2</sup> Item Reference: Item	UKATT PRS <sup>3</sup> Item Reference: Item
1. Problem focused	<p><b>Maintaining structure:</b> To what extent did the therapist attempt to structure the session?</p> <p><b>Agenda setting:</b> To what extent did the therapist articulate and implement a specific agenda for the session?</p> <p><b>Consistency of problem focus:</b> To what extent did the therapist attempt to keep the session focussed on target health behaviours?</p>	<p><b>Maintaining structure:</b> To what extent did the practitioner attempt to structure the session?</p> <p><b>Agenda setting:</b> To what extent did the practitioner articulate and implement a specific agenda for the session?</p> <p><b>Consistency of problem focus:</b> To what extent did the practitioner attempt to keep the session focussed on target problems?</p>	<p><b>Maintaining structure:</b> To what extent did the therapist attempt to structure the session?</p> <p><b>Agenda setting:</b> To what extent did the therapist articulate and implement a specific agenda for the session?</p> <p><b>Consistency of problem focus:</b> To what extent did the therapist attempt to keep the session focussed on target problems?</p>
2. Collaboration	<p><b>Open questions:</b> To what extent did the practitioner use open-ended questions to elicit greater client exploration of the target health behaviours?</p>	<p><b>Open questions:</b> To what extent did the practitioner use open-ended questions to elicit the client's perception of his/her problems, motivation, change efforts, and plans?</p>	<p><b>Therapist as an active agent of change:</b> To what extent did the therapist do things, specific tasks during the session or refer to tasks s/he had done in between sessions, on behalf of the client?</p> <p><b>Collaboration:</b> To what extent did the therapist convey that the treatment is a collaborative effort?</p>
3. Empathy	<p><b>Empathy:</b> To what extent did the practitioner respond empathically to the client, i.e. demonstrating that they have accurately understood what the client has conveyed and the meaning of this to the client?</p>	<p><b>Empathy:</b> To what extent did the practitioner respond empathically to the client, i.e. demonstrating that they have accurately understood what the client has conveyed, and the meaning of this to the client?</p>	<p><b>Empathy:</b> To what extent did the therapist respond empathically to the client, i.e. demonstrate that they have accurately understood what the client has conveyed, and the meaning of this to the client?</p>
4. Strengths and affirmation	-	<p><b>Eliciting self-efficacy for change:</b> To what extent did the practitioner attempt to elicit self-efficacy for change from the client?</p>	<p><b>Eliciting self-efficacy for change:</b> To what extent did the therapist attempt to elicit self-efficacy for change from the client?</p>

BATS Item Reference	ADAPTA Item Reference: Item	AESOPS Item Reference: Item	UKATT Item Reference: Item
5. Complex reflections	<b>Reflective listening:</b> To what extent did the practitioner reflect back what the client had said in order to communicate understanding of the client's comments and concerns?	<b>Reflective listening:</b> To what extent did the practitioner reflect back what the client had said in order to communicate understanding of the client's comments and concerns?	<b>Reflective listening:</b> To what extent did the therapist reflect back what the client had said in order to communicate understanding of the client's comments and concerns?
6. Homework assigned	<b>Prompt self-monitoring:</b> To what extent did the therapist request that the client monitor their behaviour using a worksheet or diary?	-	<b>Homework:</b> To what extent did the therapist plan or review concrete tasks assigned during therapy to be carried out outside of therapy?
7. Homework reviewed	<b>Review homework:</b> To what extent did the therapist review concrete tasks assigned as homework during the previous therapy session?	-	(Above item relevant)
8. Treatment goals	<b>Discussion of behaviour change goal:</b> To what extent did the therapist discuss or address the setting and maintaining of an appropriate behaviour change goal?	<b>Commitment to drinking goal:</b> To what extent did the practitioner discuss or address the setting and maintaining of an appropriate drinking goal?  <b>Eliciting commitment to change drinking:</b> To what extent did the practitioner elicit a commitment from the client to change their drinking?	<b>Commitment to drinking goal:</b> To what extent did the therapist discuss or address the setting and maintaining of an appropriate drinking goal?  <b>Eliciting commitment to change drinking:</b> To what extent did the therapist elicit a commitment from the client to change their drinking?
9. Developing discrepancy	-	<b>Creating conflict:</b> To what extent did the practitioner create or heighten the internal conflict or discrepancy experienced by the client between where they are currently and where they want to be?	<b>Creating conflict:</b> To what extent did the therapist create or heighten the internal conflict or discrepancy experienced by the client between where they are currently and where they want to be?

BATS Item Reference	ADAPTA Item Reference: Item	AESOPS Item Reference: Item	UKATT Item Reference: Item
10. Exploring pros and cons of change	-	<b>Eliciting optimism for change:</b> To what extent did the practitioner attempt to elicit optimism for change from the client?	<b>Eliciting optimism for change:</b> To what extent did the therapist attempt to elicit optimism for change from the client?
11. Behaviour change planning	<p><i>[Plan behaviour change: To what extent did the therapist make a concrete behaviour change plan?</i></p> <p>Item scores for ‘plan behaviour change’ were omitted from convergent validity analyses due to six missing values for frequency and quality.]</p>	-	<p><b>Alternative activities to drinking:</b> To what extent did the therapist initiate the discussion and/or planning of social activities that the client could engage in which do not involve drinking alcohol or reasons why the client might seek alternative activities to drinking?</p> <p><b>Therapist as task orientated:</b> To what extent did the therapist actively discuss specific plans designed to bring about behaviour change in the client and/or other people close to them?</p>
12. Sources of support	<p><b>Identify sources of support for change:</b> To what extent did the therapist inquire about or discuss the availability of specific individuals who will be or are sources of support for the client’s involvement in treatment or effort to change behaviour?</p> <p><b>Involvement of others in behaviour change:</b> To what extent did the therapist initiate the planning and actual involvement of other people in working towards behaviour change with the client?</p>	-	<p><b>Identify sources of support for change:</b> To what extent did the therapist inquire about or discuss the availability of specific individuals who will be or are sources of support for the client’s involvement in treatment or effort to change drinking behaviour?</p> <p><b>Involvement of others in behaviour change:</b> To what extent did the therapist initiate the planning, and actual involvement of other people in working towards behaviour change with the client?</p>

<sup>1</sup> ADAPTA PRS (11 of 15 items included) = ADAPTA (Addressing Drinking Among Patients: comparing Two Approaches) Process Rating Scale.

<sup>2</sup> AESOPS PRS (11 of 19 items included) = AESOPS (Alcohol: Evaluating Stepped care in Older Populations Study) Process Rating Scale.

<sup>3</sup> UKATT PRS (18 of 27 items included) = UKATT (United Kingdom Alcohol Treatment Trial) Process Rating Scale.

### E.13 Checks of normality for data used to examine convergent validity

The relationships between the trial-specific fidelity measures and the BATS were examined. Variables were checked for normality to determine which correlation coefficient to use. Three variables were not normally distributed (**Figure 1**); visual checks of normality were supported by the Shapiro-Wilk test. The Shapiro-Wilk test is recommended over the more popular Kolmogorov-Smirnov test for small samples (Ghasemi and Zahediasl, 2012).

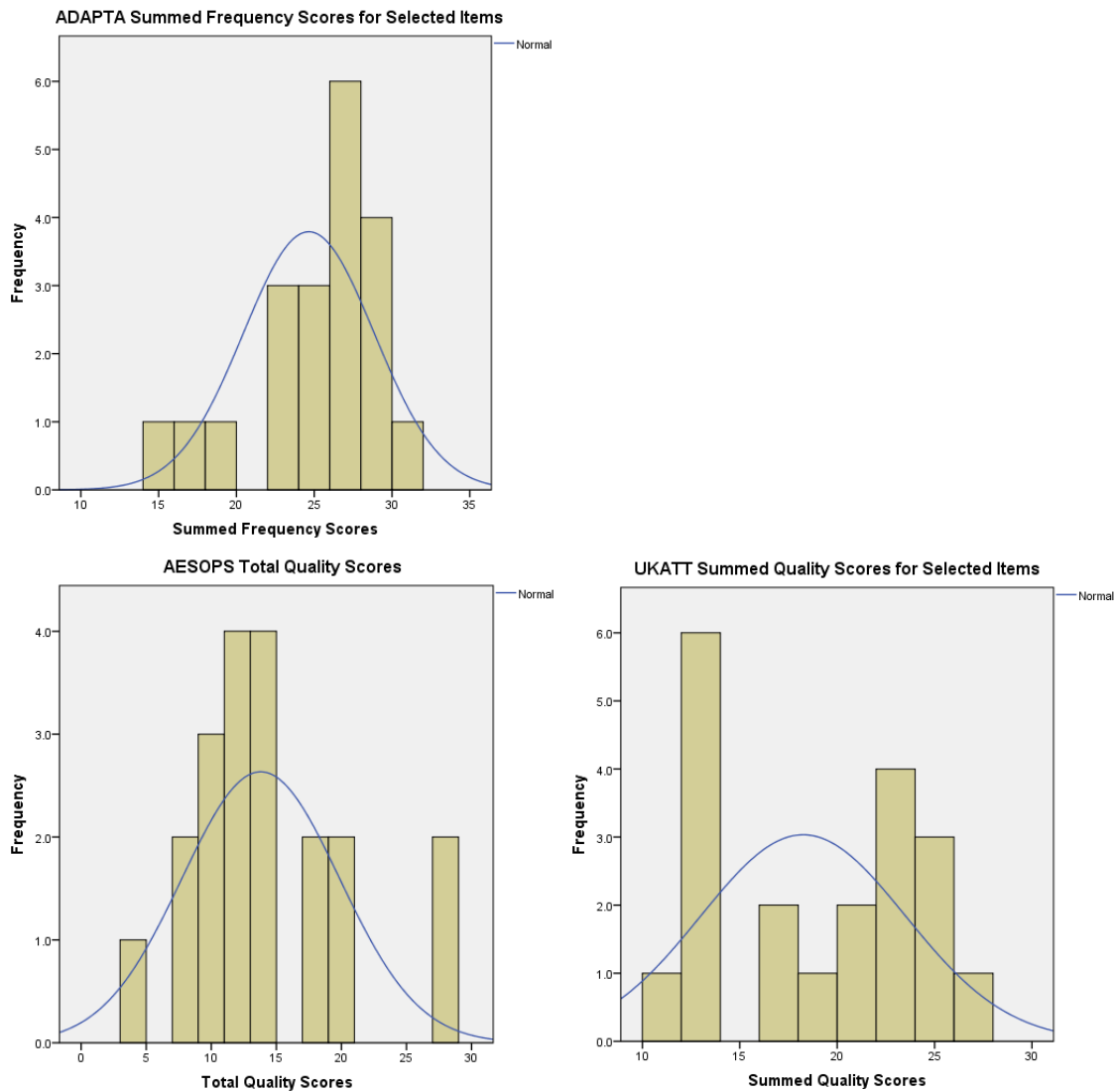


Figure 1: Example frequency distributions for convergent validity analyses

### E.14 Calculating weighted kappa for item 1 ‘problem focused’

Cohen’s kappa adjusts for the proportion of observed agreement by taking account of the amount of agreement that is expected to occur by chance (Bowers et al., 2014). The kappa coefficient ( $k$ ) is defined as:

$$k = \frac{\text{proportion of observed agreement} - \text{proportion of expected agreement}}{1 - \text{proportion of expected agreement}} = \frac{P_o - P_e}{1 - P_e}$$

Cohen’s kappa is widely used for measuring agreement for binary variables among two raters (McHugh, 2012). The kappa coefficient was generalised for ordinal data (Tang et al., 2015); known as the weighted kappa, the generalised coefficient adjusts for the degree of disagreement between response options. The magnitude of the discrepancies are taken into account by allocating weights to the different levels of disagreement (Altman, 1999; Mandrekar, 2011).

The quadratic weights were obtained by:  $w_{ij} = ((i - j)/(k - 1))^2$ , where  $i$  and  $j$  represent the rows or columns of the ratings (of the two raters) and  $k$  the maximum number of possible ratings (StataCorp, 2013a); items on the BATS were scored on a 5-point Likert scale, five was the maximum number of possible ratings. **Table 1** shows the weights for agreement, the rows and columns correspond to the ratings. An example calculation is provided below:

$$w_{ij} = 1 - \left(\frac{(i - j)}{(k - 1)}\right)^2 = 1 - \left(\frac{(1 - 2)}{(5 - 1)}\right)^2 = 1 - \left(\frac{(-1)}{(4)}\right)^2 = 1 - 0.0625 = 0.9375$$

**Table 1: Quadratic weights for the BATS ratings**

Ratings (Rater 1)	Ratings (Rater 2)					Rows ( $i$ )
	Not at all	Very little	Somewhat	A good deal	Extensively	
Not at all	<b>1</b>	0.9375	0.75	0.4375	0	<b>1</b>
Very little	0.9375	<b>1</b>	0.9375	0.75	0.4375	<b>2</b>
Somewhat	0.75	0.9375	<b>1</b>	0.9375	0.75	<b>3</b>
A good deal	0.4375	0.75	0.9375	<b>1</b>	0.9375	<b>4</b>
Extensively	0	0.4375	0.75	0.9375	<b>1</b>	<b>5</b>
Columns ( $j$ )	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	

The weighted observed and expected proportion of agreement was defined as (Altman, 1999):

$$P_{o(w)} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^k w_{ij} f_{ij} \qquad P_{e(w)} = \frac{1}{n^2} \sum_{i=1}^k \sum_{j=1}^k w_{ij} r_i c_j$$

**Table 2** and **Table 3** show the observed scores and the weighted observed scores for item 1 ‘problem focused’;  $P_{o(w)} = 0.978$ . **Table 4** and **Table 5** show the expected scores and the weighted expected scores;  $P_{e(w)} = 0.847$ .

**Table 2: Observed scores for item 1 ‘problem focused’ for two raters ( $f_{ij}$ )**

Rater 1’s Scores	Rater 2’s Scores					Total
	Not at all	Very little	Somewhat	A good deal	Extensively	
Not at all	<b>0</b>	0	0	0	0	0
Very little	1	<b>1</b>	0	0	0	2
Somewhat	0	1	<b>0</b>	0	0	1
A good deal	0	0	1	<b>1</b>	2	4
Extensively	0	0	0	2	<b>11</b>	13
Total	1	2	1	3	13	20

**Table 3: Weighted observed scores for item 1 ‘problem focused’ for two raters ( $w_{ij}f_{ij}$ )**

Rater 1’s Scores	Rater 2’s Scores				
	Not at all	Very little	Somewhat	A good deal	Extensively
Not at all	<b>0</b>	0	0	0	0
Very little	0.9375	<b>1</b>	0	0	0
Somewhat	0	0.9375	<b>0</b>	0	0
A good deal	0	0	0.9375	<b>1</b>	1.875
Extensively	0	0	0	1.875	<b>11</b>

**Table 4: Expected scores for item 1 ‘problem focused’ for two raters ( $r_i c_j$ )**

Rater 1’s Scores	Rater 2’s Scores					Total
	Not at all	Very little	Somewhat	A good deal	Extensively	
Not at all	<b>0</b>	0	0	0	0	0
Very little	2	<b>4</b>	2	6	26	2
Somewhat	1	2	<b>1</b>	3	13	1
A good deal	4	8	4	<b>12</b>	52	4
Extensively	13	26	13	39	<b>169</b>	13
Total	1	2	1	3	13	20



**Table 5: Weighted expected scores for item 1 ‘problem focused’ for two raters ( $w_{ij}r_i c_j$ )**

Rater 1’s Scores	Rater 2’s Scores				
	Not at all	Very little	Somewhat	A good deal	Extensively
Not at all	<b>0</b>	0	0	0	0
Very little	1.875	<b>4</b>	1.875	4.5	11.375
Somewhat	0.75	1.875	<b>1</b>	2.8125	9.75
A good deal	1.75	6	3.75	<b>12</b>	48.75
Extensively	0	11.375	9.75	36.5625	<b>169</b>

Weighted kappa for item 1 ‘problem focused’ was calculated as:

$$k_w = \frac{P_{o(w)} - P_{e(w)}}{1 - P_{e(w)}} = \frac{0.978 - 0.847}{1 - 0.847} = 0.857$$

### E.15 Calculating intraclass correlation coefficient for item 1 ‘problem focused’

A sample of the therapy session recordings were rated by two raters using the BATS ( $n=20$ ). Scores for item 1 ‘problem focused’ are presented in **Table 1**.

**Table 1: Scores for item 1 ‘problem focused’ for two raters**

Recording	Rater 1’s Scores*	Rater 2’s Scores*	Mean
1	4	4	4
2	4	3	3.5
3	4	4	4
4	4	4	4
5	3	2	2.5
6	4	4	4
7	3	4	3.5
8	4	4	4
9	4	4	4
10	4	4	4
11	2	1	1.5
12	3	4	3.5
13	4	4	4
14	1	0	0.5
15	4	4	4
16	1	1	1
17	4	3	3.5
18	4	4	4
19	4	4	4
20	3	3	3
<b>Mean</b>	<b>3.4</b>	<b>3.25</b>	<b>3.325</b>

\* Scores on a 5-point Likert scale: 0=not at all, 1=a little, 2=somewhat, 3=a good deal, 4=extensively.

There were three sources of variability ( $\sigma^2$ ) associated with rating the recordings. The first source is therapist (or subject) variability, represented by the differences among the mean scores for each recording (far right hand column in **Table 1**). The second source is rater variability, reflected by the differences among the mean scores for each rater (column means in **Table 1**) (Streiner et al., 2015). The last source is error variability, and reflects the difference between the true score and the observed score for each recording (not in **Table 1**). The true score for an individual recording is “the mean of an infinite number of scores” (Weir, 2005, p.232). The ICC for consistency for a single rater is defined below; rater variability was not included in the definition, as the raters were a fixed effect (Streiner et al., 2015):

$$ICC(C, 1) = \frac{\text{therapist variability}}{\text{therapist variability} + \text{error variability}} = \frac{\sigma_{therapists}^2}{\sigma_{therapists}^2 + \sigma_{error}^2}$$

ICC estimates can be derived from a repeated measures Analysis of Variance (ANOVA) (Streiner et al., 2015). **Table 2** gives the ANOVA summary table for item 1 ‘problem focused’ data. The definition of the ICC can, therefore, be written as:

$$ICC(C, 1) = \frac{MS_{therapists} - MS_{error}}{MS_{therapists} + (k - 1)MS_{error}}$$

**Table 2: Analysis of variance summary table for item 1 ‘problem focused’**

Source of variation	Sum of squares	Degrees of freedom	Mean square	F	Significance
Model	487.725	21	23.225	134.740	0.000
Therapists	45.275*	19	2.383	13.824	0.000
Raters	0.225**	1	0.225	1.305	0.267
Error	3.275	19	0.172		
Total	491.000	40			

\*Sum of squares (therapists) =  $3 \times [(4 - 3.325)^2 + (3.5 - 3.325)^2 + \dots + (3 - 3.325)^2] = 45.275$

\*\*Sum of squares (raters) =  $20 \times [(3.4 - 3.325)^2 + (3.25 - 3.325)^2] = 0.225$

The mean square (MS) values from **Table 2** relate to the variances due to therapists and error:

- $MS_{therapists} = 2\sigma_{therapists}^2 + \sigma_{error}^2 = 2.383$
- $MS_{error} = \sigma_{error}^2 = 0.172$

The equations above can be manipulated to calculate the variances:

- $\sigma_{therapists}^2 = \frac{MS(therapists) - MS(error)}{2} = \frac{2.383 - 0.172}{2} = 1.1055$
- $\sigma_{error}^2 = MS_{error} = 0.172$

The variances can be then used to calculate the ICC for item 1 ‘problem focused’:

- $ICC(C, 1) = \frac{\sigma_{therapists}^2}{\sigma_{therapists}^2 + \sigma_{error}^2} = \frac{1.1055}{1.1055 + 0.172} = 0.87$

The reliability estimate of 0.87 reflects the degree of consistency for a single rater. Had the raters been randomly selected, the degree of absolute agreement would have been assessed and rater variability incorporated. Absolute agreement concerns the degree that scores provided by raters are similar in absolute value (Hallgren, 2012), i.e. whether raters assigned the same rating for the same therapist recording, and is defined as:

$$ICC(A, 1) = \frac{\sigma_{therapists}^2}{\sigma_{therapists}^2 + \sigma_{raters}^2 + \sigma_{error}^2}$$

The reliability estimate for item 1 'problem focused' would be calculated as:

$$ICC(A, 1) = \frac{MS_{therapists} - MS_{error}}{MS_{therapists} + \frac{k}{n}(MS_{raters} - MS_{error}) + (k - 1)MS_{error}}$$

Where  $k$  is the number of raters, and  $n$  the number of recordings; therefore:

$$ICC(A, 1) = \frac{2.383 - 0.172}{2.383 + (0.1 \times (0.225 - 0.172)) + (1 \times 0.172)} = \frac{2.211}{2.383 + 0.0053 + 0.172} = 0.86$$

The reliability estimate of 0.86 reflects the degree of absolute agreement for a single rater; the estimate is equivalent to the weighted kappa.

### E.16 Sample size estimation for the ICC

The number of recordings required to estimate inter-rater reliability can be calculated using the confidence interval (CI) of the ICC. A hypothetical study is given as an example to show how the sample size can be calculated. A reliability study is being conducted with two raters. The estimated effect size is 0.80 ( $R$ ), and the maximum acceptable width of the 95% CI is 0.2; that is, there is a 95% chance that the true ICC would lie between 0.60 (0.80-0.2) and 1.00 (0.8+0.2). The standard error<sup>92</sup> ( $SE$ ) is half of the CI, 0.1. The required sample size in this example study was calculated using the steps described by Streiner et al. (2015):

1. Compute  $R^-$ , a more conservative estimate than  $R$ :

$$R^- = R - SE = 0.8 - 0.1 = 0.7$$

2. Compute the log transformed values of  $R$  and  $R^-$ . The Fisher ( $z_R$ ) transformation is used to remove the skewness in the standard error ( $k$  = the number of observations per recording,  $R$  = the estimated ICC):

$$z_R = \frac{1}{2} \log_e \left[ \frac{1 + (k - 1)R}{1 - R} \right] = \frac{1}{2} \log_e \left[ \frac{1 + (1 \times 0.8)}{1 - 0.8} \right] = \frac{1}{2} \log_e \left[ \frac{1.8}{0.2} \right] = 1.099$$

$$z_{R^-} = \frac{1}{2} \log_e \left[ \frac{1 + (k - 1)R^-}{1 - R^-} \right] = \frac{1}{2} \log_e \left[ \frac{1 + (1 \times 0.7)}{1 - 0.7} \right] = \frac{1}{2} \log_e \left[ \frac{1.7}{0.3} \right] = 0.867$$

3. Compute the SE of the z-scores:

$$SE = z_R - z_{R^-} = 1.099 - 0.867 = 0.232$$

4. Compute the required sample size ( $n$ ):

$$n = 2 + \frac{k}{2(k - 1)(z_R - z_{R^-})^2} = 2 + \frac{2}{2(2 - 1)(0.232)^2} = 2 + \frac{2}{2 \times 1 \times (0.232)^2} = 21$$

Therefore, the study would need 21 recordings to compute a reliability estimate of 0.80 with a standard error of 0.1.

---

<sup>92</sup> The sample ICC is an estimate of the population ICC, the standard error is the standard deviation of the sampling distribution of the ICC (Bland, 2017).

