

Assessing maladaptive overcontrol:
Development and validation of the
Overcontrol Assessment Questionnaire

Aleksandra E. Lambert

*A thesis submitted in partial fulfilment of the
requirements of Nottingham Trent University for the
degree of Doctor of Philosophy*

August 2024

Copyright Statement

The copyright in the intellectual property of the author. You may copy up to 5% of this work for private study or personal, non-commercial research. Any re-use of the information contained within this document should be fully referenced, quoting the author, title, university, degree level, and pagination. Queries or requests for any other use or for a more substantial copy are required and should be directed to the author.

Table of Contents

Acknowledgements	13
Abstract.....	15
Dissemination of findings.....	17
Abbreviations	19
1. Introduction	21
1.1 Background	21
1.2 Research aims and objectives.....	23
1.3 Overview of the chapters	24
2. Literature review	27
2.1 The construct of self-control	27
2.1.1 <i>Self-control and self-regulation</i>	29
2.1.2 <i>The components of self-control</i>	30
2.1.8 <i>The current thesis' definition of self-control</i>	33
2.2 Linear view of self-control.....	33
2.2.1 <i>The strength model of self-control</i>	34
2.2.2 <i>The General Theory of Crime</i>	38
2.4 Quadratic view of self-control	40
2.4.1 <i>Megargee's idea of under- and overcontrolled offending</i>	40
2.4.2 <i>Ego-control and ego-resilience</i>	42
2.5 The Neurobiosocial Theory for Disorders of Overcontrol.....	46
2.5.1 <i>An overview of Lynch's theoretical framework</i>	47
2.5.2 <i>The Neuroregulatory Model of Socioemotional Functioning</i>	49
2.5.3 <i>The Biosocial Theory for Disorders of Overcontrol</i>	50
2.5.4 <i>Integrating Lynch's neuroregulatory and biosocial models</i>	54
2.5.4 <i>Maladaptive overcontrol and links to psychopathology</i>	54
2.5.6 <i>Radically Open Dialectical Behaviour Therapy (RO DBT)</i>	55
2.5.7 <i>The evidence gap</i>	57
2.6 The current protocol for assessing maladaptive overcontrol	60
2.6.1 <i>Step 1 – the self-report questionnaires</i>	60
2.6.2 <i>Step 2 – the diagnostic interview</i>	63
2.6.3 <i>Step 3 – clinician-rated questionnaires</i>	63
2.6.4 <i>The need for a better psychometric measure</i>	65
2.7 Rationale	65
3. Conceptual framework, functional properties, and item generation	68
3.1 Introduction.....	68
3.1.1 <i>Conceptual framework</i>	69
3.2.2 <i>Functional properties of the scale</i>	70
3.2.1 <i>Scaling model</i>	71
3.2.2 <i>Mid-point response option</i>	72
3.2.3 <i>Number and labelling of response options</i>	73
3.3 Item development.....	74
3.4 Deductive item generation	75
3.5 Inductive item generation.....	75
3.5.1 <i>Pilot study</i>	75
3.5.1 <i>Main study</i>	81
3.6 Data collation and results	87
3.7 Discussion	91
4. Content validation via expert judgement method.....	95
4.1 Introduction	95
4.1.1 <i>The expert judges</i>	96
4.1.2 <i>Approaches for establishing content validity via expert judgement</i>	96
4.1.3 <i>Methodological considerations</i>	98
4.1.4 <i>The current study</i>	99
4.2 Method	100
4.2.1 <i>Design</i>	100
<i>Follow-up review round</i>	100
4.2.2 <i>Participants</i>	101
4.2.3 <i>Materials</i>	103

4.2.4 Procedure.....	104
4.3 Data collation and results	105
4.3.1 Initial review	105
4.3.2 Follow-up review	109
4.3.3 The new item pool	109
4.4 Discussion	112
5. Cognitive interviews with the target population	115
5.1 Introduction.....	115
5.1.1 Pre-testing the scale.....	115
5.1.2 Content validation via target population judgement	116
5.1.3 Cognitive interviewing	117
5.1.3 The current study.....	119
5.2 Method	120
5.2.1 Design.	120
5.2.2 Participants.	120
5.2.3 Materials	123
5.2.4 Procedure.....	124
5.3 Data collation and results	126
5.4 Discussion and conclusions.....	135
6. Internal structure and reliability of the OAQ.....	139
6.1 Introduction	139
6.1.1 Exploratory factor analysis (EFA).....	140
6.1.2 Confirmatory factor analysis (CFA).....	142
6.1.4 Methodological considerations for conducting factor analyses.....	144
6.1.5 Assessing internal consistency of psychometric scales.....	146
6.1.6 The current studies.....	148
6.2 Method	149
6.2.1 Design.	149
6.2.2 Participants.	150
6.2.3 Materials.	151
6.2.4 Procedure.....	151
6.3 Data collation	152
6.3.1 Data cleaning.....	152
6.3.2 Missing data	153
6.3.3 Data split.....	153
6.4 Demographic characteristics: Data analysis and results	154
6.5 Exploratory Factor Analysis (EFA): Data analysis and results	155
6.5.1 The low flexible control (LFC) domain.....	156
6.5.2 The low receptivity and openness (LRO) domain.....	158
6.5.3 The low social connectedness and intimacy with others (LSC) domain.....	160
6.5.4 The pervasive inhibited emotional expression and low emotional awareness (PIE) domain.....	164
6.5.4 The global EFA model.....	168
6.6 Confirmatory Factor Analysis (CFA): Data analysis and results.....	174
6.7 Discussion	178
6.7.1 Outcomes of the studies.....	178
6.7.2 Strengths and limitations.....	182
6.7.3 Conclusions.....	183
7. Locating maladaptive overcontrol within the wider literature and further psychometric evaluation the OAQ.....	185
7.1 Introduction	185
7.1.1 Relationships between constructs and the scale's validity.....	185
7.1.2 Maladaptive overcontrol and the links to other constructs.....	188
7.1.3 Test-retest reliability.....	195
7.1.4 Aims and hypotheses.....	196
7.2 Methods.....	197
6.2.1 Design.....	197
Participants.....	197
7.2.3 Measures.....	199
7.2.4 Procedure.....	203

7.3 Data collation and analysis.....	204
7.3.1 <i>Maladaptive overcontrol in relation to other constructs.</i>	204
7.3.2 <i>Test-retest reliability.</i>	205
7.4 Results	206
7.4.1 <i>Maladaptive overcontrol and other psychological constructs.</i>	206
7.4.2 <i>Test-retest reliability.</i>	213
7.5 Discussion.	214
7.5.1 <i>The OAQ: Internal consistency, test-retest reliability, and inter-factor correlations.</i> ..	214
7.5.2 <i>Maladaptive overcontrol and ego-control.</i>	215
7.5.3 <i>Maladaptive overcontrol and ego-resilience.</i>	217
7.5.4 <i>Maladaptive overcontrol and behavioural inhibition.</i>	218
7.5.5 <i>Maladaptive overcontrol and behavioural activation.</i>	219
7.5.6 <i>Maladaptive overcontrol and the dimensions of impulsive behaviour.</i>	220
7.5.7 <i>Maladaptive overcontrol and the linear view of self-control.</i>	222
7.5.8 <i>Conclusion.</i>	227
8. Conclusions and reflections.....	229
8.1 Synthesis of study methods and results.....	229
8.1.1 <i>Conceptual framework.</i>	230
8.1.2 <i>Item generation.</i>	230
8.1.3 <i>Content validation.</i>	231
8.1.4 <i>Pre-testing of the scale.</i>	232
8.1.5 <i>Factor analyses.</i>	232
8.1.6 <i>Psychometric properties of the OAQ and the links between maladaptive overcontrol and</i> <i>other psychological constructs.</i>	234
8.1.7 <i>Research outcomes.</i>	235
8.2 Sample considerations.....	235
8.2.1 <i>General population and cultural and linguistic considerations.</i>	236
8.2.3 <i>Prolific recruitment.</i>	238
8.3 Embracing the iterative approach to scale and theory development.	240
8.4 Navigating the complexities of psychometric research.	242
8.5 Conclusion.....	243
References	245
Appendices	285

List of Tables

Table 3.1: The conceptual framework of maladaptive overcontrol.

Table 3.2: ASC-WP scores of participants who completed the pilot screening survey.

Table 3.3: ASC-WP scores of participants in the pilot focus group.

Table 3.4: ASC-WP scores of participants who completed the main screening survey.

Table 3.5: ASC-WP scores of post-pilot focus group participants.

Table 3.6: The initial item pool for the LFC domain of maladaptive overcontrol.

Table 3.7: The initial item pool for the LRO domain of maladaptive overcontrol.

Table 3.8: The initial item pool for the LSC domain of maladaptive overcontrol.

Table 3.9: The initial item pool for the PIE domain of maladaptive overcontrol.

Table 4.1: Experience and expertise of the experts involved in the study.

Table 4.2: Definitions and indicators for the expert rating criteria.

Table 4.3: Changes made to items that did not meet the CVI cut-off for the Coherence or Clarity criteria.

Table 4.4: Items to which semantic changes were made based on expert feedback.

Table 4.5: Items modified based on expert feedback indicating a better fit to a different domain based on domain definition.

Table 4.6: Items to which changes were applied due to potential correlation with constructs such as emotionality and affect.

Table 4.7: Decisions regarding retention and rejection of the follow-up items that did not meet the CVI cut-off for the Relevance criterion.

Table 4.8: Refined item pool following two rounds of expert review.

Table 4.9: Domain and combined Ave-CVI for Relevance, Coherence, and Clarity criteria after the follow-up review.

Table 5.1: Changes made to the items following cognitive interviews with the potential respondents.

Table 5.2: Items removed from the item pool based on cognitive interviews with the potential respondents.

Table 5.3: Items from which absolute, frequency, or intensity words have been removed following cognitive interviews with the potential respondents.

Table 5.4: Refined item pool following target population cognitive interviews.

Table 6.1: Differences in age between EFA and CFA participant samples following sample split.

Table 6.2: Differences in selected demographic characteristics (categorical) between EFA and CFA participant samples following sample split.

Table 6.3: Items in the LFC domain that were rejected based on consequent EFA models.

Table 6.4: Items retained in the LFC domain following lower-order factor EFA.

Table 6.5: Items in the LRO domain that were rejected based on consequent EFA models.

Table 6.6: Items retained in the LRO domain following lower-order factor EFA.

Table 6.7: Items in the LSC domain rejected based on the one-factor EFA model.

Table 6.8: Items in the LSC domain rejected based on the two-factor EFA model.

Table 6.9: Factor loadings of items in the LSC domain indicated by a two-factor EFA.

Table 6.10: Items retained in the LSC domain following lower-order factor EFA.

Table 6.11: Items in the PIE domain rejected based on the one-factor EFA model.

Table 6.12: Factor loadings of items in the PIE domain indicated by a two-factor EFA.

Table 6.13: Items retained in the PIE domain following lower-order factor EFA.

Table 6.14: Factor loadings of items indicated by a five-factor global EFA.

Table 6.15: Factor loading of items in a final, four-factor global EFA model.

Table 6.17: Standardised loadings of items retained following the refinement of the item pool based on the higher- order CFA results.

Table 7.1: Descriptive statistics for the scale and subscale scores of the participants who took part in the first part of the study.

Table 7.2: Inter-factor correlations within the OAQ.

Table 7.3: Correlations between the OAQ and the EUC-13 scales.

Table 7.4: Correlations between the OAQ and the BAI-BAS subscales.

Table 7.5: Correlations between the OAQ and the S-UPPS-P subscales.

Table 7.6: Correlations between the OAQ and the LSCS scale and its subscales.

Table 7.7: Descriptive statistics for the OAQ global and subscale scores of the participants who participated in both the validation and the test-retest reliability parts of the study.

List of Figures

Figure 1.1: Outline and structure of the thesis.

Figure 6.1: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the LFC domain.

Figure 6.2: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the LRO domain.

Figure 6.3: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the LSC domain.

Figure 6.4: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the LSC domain following the rejection of two items with factor loadings $<.40$.

Figure 6.5: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the LSC domain after the rejection of Factor 1.

Figure 6.6: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the PIE domain.

Figure 6.6: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the PIE domain following removal of Factor 2 items.

Figure 6.7: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the first global EFA model.

Figure 6.8: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the second global EFA model.

List of Appendices

Appendix A: A pilot focus group schedule for the item generation study.

Appendix B: Demographic information about participants in the pilot focus group.

Appendix C: Demographic information about participants in the main study focus groups.

Appendix D: Items rejected based on the I-CVI values $< .80$ for the Relevance criterion.

Appendix E: Demographic characteristics (categorical) of the participant samples recruited for the cognitive interviews study in the first round.

Appendix F: Demographic characteristics (categorical) of the participant samples recruited for the cognitive interviews study in the second round.

Appendix G: Demographic characteristics (categorical) of the total sample of participants recruited for the EFA and CFA studies.

Appendix H: Differences in selected demographic characteristics (categorical) between EFA and CFA participant samples following sample split.

Appendix I: Skewness and kurtosis values for scales (and scale subscales) included in the battery of measures.

Appendix J: Demographic characteristics (categorical) of the sample used to validate the OAQ.

Appendix K: Demographic characteristics (categorical) of the sample used to examine the test-retest validity of the OAQ.

Appendix L: Correlation matrix of measures included in the battery of measures.

Appendix M: The final version of the OAQ with participant and scoring instructions.

Acknowledgements

I dedicate this thesis to Edward. Thank you for the encouragement and warmth you offered me when I doubted myself, and for your patience and understanding when research took priority. Only you truly know what a wonderful yet frightening rollercoaster these past three years have been, and I am truly grateful for the chance to complete this ride with you by my side.

Thank you to my wonderful supervisory team – Belinda, Christine, Laura, and Pedro. You have been incredibly supportive of me every step of the way; thanks to that, I have grown both as an individual and as a professional. It was great working with you.

Thank you to Frederica. You have been my rock (or, as you say, my pebble), celebrating my highs and lifting me up in my lows. I know I have found a friend for life, no matter how many miles may separate us.

Thank you to Anna and Filip, for wholeheartedly believing in my skill and my determination; for laughing with me on the better days and helping me through the more difficult ones. I am honoured to have you as friends.

Thank you to my Mum, for cheering me on when I decided to move to a different country to pursue my dreams. I would not be at this point in my life without you. I am so, so grateful for you, and I miss you every day.

Thank you to Adrienne, Sandra, and Roger – your incessant support truly means the world to me. Not once have you doubted me, not even when I doubted myself.

Thank you to the inspiring network of academics and clinicians that I had the pleasure to network, collaborate, and share my passion for research with. A special thank you to the expert judges – Dr Ellen Astrachan-Fletcher, Dr Lee Anna Clark, Dr Ata Ghaderi, Dr Julianna Gorder, Dr Karyn Hall, Dr Roelie Hempel, and Dr Nicole Little – for their invaluable contributions to the development of the scale.

And lastly, thank you to all the people whose participation in my studies made this PhD possible. Your willingness to share your experiences and insights has been invaluable to this research.

Abstract

The idea of excessive self-control – or maladaptive overcontrol – has been present in psychological literature for several decades, including in clinical, forensic, and personality domains of psychology. However, there has been a lack of a comprehensive and testable theoretical framework that would both demonstrate how maladaptive overcontrol may present among populations and account for the mechanisms behind the development, reinforcement, and maintenance of maladaptive overcontrolled tendencies. This resulted in a linear view of self-control – i.e., the more self-control, the better – becoming the status quo among both the social science professionals and the general population. The topic of overcontrol was becoming less and less popular over time despite indications that too much self-control can be as problematic for individuals as too little self-control and that it requires tailored interventions – ultimately leading to the overcontrolled issues being overlooked and the individuals struggling with maladaptive overcontrol unable to access appropriate treatment.

In 2018, Thomas Lynch published a novel, comprehensive theoretical framework to explain maladaptive overcontrol – the Neurobiosocial Theory for Disorders of Overcontrol. The framework was developed based on nearly three decades of interdisciplinary empirical research. It integrated top-down and bottom-up neuroregulatory models of socioemotional functioning and built on those by theorising how the interactions between biological predispositions, environmental factors, and coping mechanisms function to develop, reinforce, and maintain specific deficits posited to be associated with maladaptive overcontrol. Radically Open Dialectical Behaviour Therapy was also developed, designed specifically to target these overcontrolled deficits, with preliminary evaluation studies indicating good outcomes.

Still, a valid and reliable instrument that would allow researchers to efficiently and confidently identify maladaptive overcontrol in the light of Lynch's theorising was lacking. The diagnostic protocol published alongside the theory required input from a specially trained clinician, was rather lengthy, and some of its elements were semantically complex and not well validated. This was considerably hindering research progress, making theory testing costly and difficult and slowing down the roll-out of the accompanying therapeutic intervention on a wider scale.

The current programme of research aimed to answer this urgent need for a valid and reliable questionnaire that would allow researchers to assess adults for maladaptive overcontrol. Rigorous psychometric techniques guided by an extensive review of psychometric literature were used to achieve this aim. The scale development process was theory-driven, with the conceptual framework derived directly from the Neurobiosocial Theory for Disorders of Overcontrol. Deductive and inductive methods were used to generate the initial item pool, the contents of which were then refined and validated using the expert judgement method. The item pool was then pre-tested with a sample of participants from the general population using cognitive interviews. Subsequently, two factor-analytic studies were conducted to guide decisions on factor retention and item reduction and test model fit.

As a result, a 26-item, self-report Overcontrol Assessment Questionnaire (OAQ) was developed. The four-dimensional scale was designed not only to assess the overall extent of issues associated with maladaptive overcontrol, but also its specific deficits. The OAQ presented with satisfactory internal consistency and composite reliability, as well as 4-week test-retest reliability.

The relationships between maladaptive overcontrol, as measured by the OAQ, and other constructs posited to be conceptually linked to maladaptive overcontrol were also tested, and a complex landscape of correlations was revealed. Highly overcontrolled individuals were found to present with high behavioural inhibition and low ego-control in support of Lynch's framework. More nuanced relationships were indicated between the deficits of maladaptive overcontrol and various elements of behavioural activation and impulsive behaviour. Maladaptive overcontrol has shown some convergence with ego-overcontrol, however, unexpected patterns of correlations were revealed between Lynch's model of maladaptive overcontrol and other, linear models of self-control. These mixed results warrant further investigation.

The current programme of research involved the development of a new self-report scale to assess adults for maladaptive overcontrol that presented with encouraging preliminary validity and reliability evidence, thereby offering an original contribution to psychological science. It is hoped that future research will employ the scale in empirical studies aiming to better understand the nature of maladaptive overcontrol, as well as validate the questionnaire for use in clinical and forensic populations and introduce linguistic and cultural adaptations.

Dissemination of findings

Lambert, A., Winder, B., Norman, C., Hamilton, L., Pires, P. (2024, August). Assessing maladaptive overcontrol: Development and validation of the Overcontrol Assessment Questionnaire [Poster presentation]. ECP21 Conference, Berlin, Germany.

Lambert, A., Winder, B., Norman, C., Hamilton, L., Pires, P. (2024, April). Assessing maladaptive overcontrol: Development and validation of a new self-report Overcontrol Assessment Questionnaire [Conference presentation]. BSPiD14 Conference, Salford, Manchester, UK.

Lambert, A., Winder, B., Norman, C., Hamilton, L., Pires, P. (2024, March). Factor structure and internal consistency reliability of the Overcontrol Assessment Questionnaire [Guest talk]. RO DBT Research Committee Meeting, Online.

Lambert, A., Winder, B., Norman, C., Hamilton, L., Pires, P. (2023, December). Assessing maladaptive overcontrol: Development and validation of a new self-report measure [Conference presentation]. APE 2023 Conference, Online.

Lambert, A., Winder, B., Norman, C., Hamilton, L., Pires, P. (2023, June). Assessing maladaptive overcontrol: Development and validation of a new self-report measure [Conference presentation]. RO DBT Conference 2023: Better Together, Evanston, Illinois, USA.

Lambert, A., Winder, B., Norman, C., Hamilton, L., Pires, P. (2023, April). Assessing maladaptive overcontrol: Development and validation of a new self-report measure [Poster presentation]. Strategic Research Themes Conference, Nottingham, UK.

Abbreviations

AAQ-II	Acceptance and Action Questionnaire-II (Bond et al., 2011)
ASC-WP	Assessing Styles of Coping: Word-Pair Questionnaire (Lynch, 2018)
Ave-CVI	Average Content Validity Index
BAI	Behavioral Inhibition and Activation Scales (Carver & White, 1996)
BAS	Behavioral Activation Scale (Carver & White, 1996)
BIS	Behavioral Inhibition Scale (Carver & White, 1996)
BSCS	Brief Self-Control Scale (Tangney et al., 2004)
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CVI	Content Validity Index
CVR	Content Validity Ratio
EFA	Exploratory Factor Analysis
ER-89	Ego Resilience Scale (Block & Kremen, 1996)
EUC-13	Ego Undercontrol Scale-13 (Isaksson, Ghaderi, Wolf-Arehult, & Ramklint et al., 2021)
I-CVI	Item Content Validity Index
LFC	Low Flexible Control
LRO	Low Receptivity and Openness
LSC	Low Social Connectedness
LSCS	Low Self-Control Scale (Grasmick et al., 1993)
MINRES	Minimum Residuals
OAQ	Overcontrol Assessment Questionnaire
PIE	Pervasive Inhibited Emotional Expression and Low Emotional Awareness
PNS	Personal Need for Structure (Neuberg & Newsom, 1993)
RMSEA	Root Mean Square Error of Approximation
RO DBT	Radically Open Dialectical Behavior Therapy
RST	Reinforcement Sensitivity Theory
RUO	Resilient-Undercontrolled-Overcontrolled Framework
S-UPPS-P	Short UPPS-P ([Negative] Urgency, [Lack of] Premeditation, [Lack of] Perseverance, Sensation Seeking, [Positive] Urgency) Impulsive Behavior Scale (Lynam, 2013)
SRMR	Standardised Root Mean Square Residual
UA-CVI	Universal Agreement Content Validity Index
UK	United Kingdom
USA	United States of America
TLI	Tucker-Lewis Index
WEIRD	Western, Educated, Industrialised, Rich, and Democratic
WLSMV	Weighted Least Squares with Means and Variance Adjusted

1. Introduction

1.1 Background

Over the years, numerous studies have linked low self-control to a wide range of negative life outcomes – ranging from poor grades at school and less financial stability, to complex psychopathology and offending behaviour (e.g., Moffitt et al., 2011; Tangney et al., 2004). As a result, the ability to exert self-control was portrayed as “*a hallmark of adaptation*” (De Ridder et al., 2012, p. 76), and the belief that more self-control invariably equates to better outcomes – i.e., a linear view of self-control – became the status quo within the social sciences (Hamilton, 2021). Meanwhile, the idea of excessive self-control – maladaptive overcontrol – has been present in psychological literature for several decades (e.g., Block & Block, 1980; Megargee, 1966). Yet, with the linear view of self-control being widely endorsed, the concept of overcontrol lacked prominence and failed to gain traction.

In 2018, a novel theory was published that attempted to challenge the status quo – called the Neurobiosocial Theory for Disorders of Overcontrol. The theory developer, Thomas Lynch, asserted that adaptive self-control was not *high* self-control, but rather *flexible* self-control. He stated that each person leans towards under- or overcontrol – but most people possess the ability to flexibly adapt to changing circumstances and exert or relinquish self-control as required. However, he proposed that for those who are not able to respond flexibly, both insufficient self-control (maladaptive undercontrol) and excessive self-control (maladaptive overcontrol) could lead to adverse outcomes. The theoretical model attempted to explain how specific temperamental and socio-emotional factors could result in the development and maintenance of maladaptive overcontrol, as well as theorised how it may manifest among individuals. Lynch conceptualised maladaptive overcontrol as a multidimensional construct that manifests through four core deficits that negatively impact individual’s mental health and well-being: low flexible control, low receptivity and openness, low social connectedness and intimacy with others, and pervasive inhibited emotional expression and low emotional awareness.

Lynch (2018) also developed a novel transdiagnostic therapeutic intervention called Radically Open Dialectical Behaviour Therapy (RO DBT). It was designed

specifically to address the deficits Lynch identified as characteristic of maladaptive overcontrol, rather than focus on targeting symptoms of disorders. Preliminary evidence from early evaluation studies indicated that the intervention can be helpful in addressing problematic overcontrolled behaviours in clients with diagnoses of treatment-resistant depression (Gilbert et al., 2023; Lynch et al., 2018), anorexia nervosa (e.g., Lynch et al., 2013; Isaksson, Ghaderi, Ramklint, et al., 2021; Isaksson, Ghaderi, Wolf-Arehult, Öster, & Ramklint, 2021), perfectionism (Little & Codd III, 2020), and autism spectrum conditions (Cornwall et al., 2021). Because RO DBT was directly based on the theoretical model, the promising results of studies piloting the intervention also provided indirect evidence in support of the model. Nevertheless, the Neurobiosocial Theory for Disorders of Overcontrol is yet to undergo thorough empirical testing that would validate the multidimensional nature of overcontrol and the claims regarding the development, maintenance, and clinical presentations of maladaptive overcontrolled personality.

Establishing a robust evidence base for the theoretical model is crucial, as literature suggests that overcontrol is present and prevalent among various populations (Hamilton, 2021; Lynch, 2018). Many of the commonly offered interventions may not be effective in treating issues of overcontrol due to treatment targets not being aligned with the specific needs of overcontrolled individuals (Davey et al., 2005; Hamilton, 2021). There is a need to recognise that (1) not only insufficient but also excessive self-control can lead to psychopathology, (2) changes in self-control tendencies can serve as a mechanism of psychological change, and (3) treatments tailored to one's self-control tendencies need to be introduced on a wider scale (Vanderbleek & Gilbert, 2018).

However, the research (and clinical) progress in the area is being hindered by the lack of a validated, resource-efficient tool that allows for a prompt and accurate identification of maladaptive overcontrol, and yet paints a comprehensive picture of the difficulties that overcontrolled individuals experience. While thorough, the current diagnostic protocol (Lynch, 2018) requires input from a specially trained clinician, is time-consuming to complete, and there is no evidence for its validity. As such, there is an urgent need for a new assessment tool, so that the theoretical model can be

thoroughly tested and further developed, and so that eventually, tailored treatment can be offered on a wider scale to overcontrolled individuals.

1.2 Research aims and objectives

The overarching aim of the current programme of work was to answer the need for a robust measure of maladaptive overcontrol. This thesis presents the psychometric process of developing and validating a new, multidimensional, self-report measure of maladaptive overcontrol – the Overcontrol Assessment Questionnaire (OAQ). The objectives of the project were as follows:

1. To develop an initial item pool for the new measure using a combination of deductive and inductive methods for item generation.
2. To evaluate the content validity of the item pool using the expert judgement method and refine the item pool accordingly.
3. To conduct cognitive interviews with the potential respondents to further refine the item pool.
4. To evaluate the internal structure of the newly developed measure using exploratory and confirmatory factor analyses and appropriate internal consistency indices.
5. To further evaluate the internal properties of the scale, including the test-retest reliability, and place the construct of maladaptive overcontrol within the wider literature relative to other, potentially related psychological constructs.

A theory-driven approach guided the psychometric process throughout, with the Neurobiosocial Theory for Disorders of Overcontrol (Lynch, 2018) being the pivotal theoretical framework. The theory, albeit yet to be empirically validated, offers a clear and coherent framework for the construct of overcontrol based on nearly three decades of translational research. As such, a priori approach is an appropriate choice for the development of the current measure that would allow for a systematic process for psychometric development despite limited empirical evidence and an absence of other validated measures pertaining to the latent construct. Moreover, analyses of the internal structure of a carefully developed scale based on a specific theoretical model allows for a preliminary examination of the theorised structure of the latent construct.

What follows is an overview of the chapter contents. A pictorial representation of the structure of the thesis is presented in Figure 1.1.

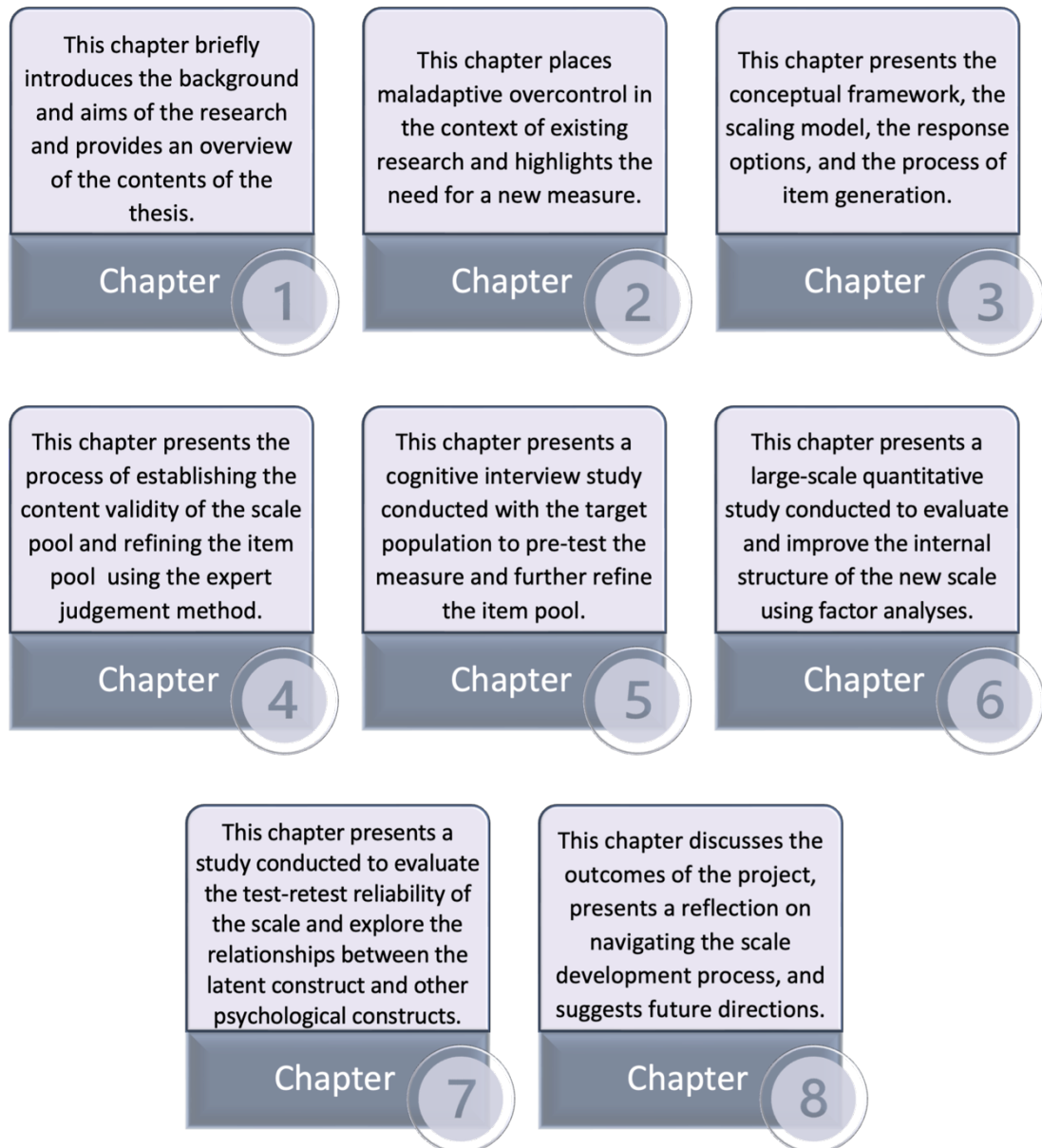


Figure 1.1: Outline and structure of the thesis.

1.3 Overview of the chapters

Chapter 2 discusses existing research relevant to the concept of overcontrol. Firstly, various definitions of self-control are considered, followed by a review of previous literature pertaining to overcontrol, undercontrol, and self-control psychopathology. A detailed overview of the Neurobiosocial Theory for Disorders of Overcontrol underpinning the assessment tool is also provided. Finally, existing assessments of

self-control and overcontrol are discussed, culminating in a rationale highlighting the importance of the current programme of work.

Chapter 3 places the construct of maladaptive overcontrol within a conceptual framework developed in accordance with the Neurobiosocial Theory for Disorders of Overcontrol. It establishes the scaling model and response options for the OAQ and presents the process of generating the initial item pool using deductive and inductive methods of item generation used in the development of the scale.

Chapter 4 discusses the expert judgement method of content validation and considers methodological issues and recommendations around the approach. It presents a two-round content validity study in which expert judges familiar with the underlying theoretical framework rate the scale items and domains according to carefully developed rating criteria and provide qualitative feedback for improvement.

Chapter 5 focuses on the use of cognitive interviews in the psychometric scale development process. It discusses the utility and process of cognitive interviewing and presents a study which utilised online cognitive interviews with the target population to further refine the item pool for the OAQ.

Chapter 6 discusses the process of evaluating and refining internal structures of psychometric measures. It presents a large-scale quantitative study which utilised Exploratory and Confirmatory Factor Analyses to systematically evaluate and improve the internal structure of the newly developed scale as means of ensuring its construct validity. Internal consistency analyses are also presented.

Chapter 7 discusses the purpose psychometrics scales serve in theory building and the importance of the stability of scale scores over time. The chapter presents a study aiming to test some of the assumptions of the theoretical model through exploring the relationships between maladaptive overcontrol and other, potentially linked psychological constructs, as well explore the test-retest reliability of the OAQ.

Chapter 8 brings together the previous chapters and offers a summary of the scale development process, its outcomes, and implications. A critical reflection on navigating the psychometric process is also provided, and future directions for research are suggested. A final version of the OAQ is presented, alongside instructions, scoring procedures, and recommendations for the interpretation of scores.

2. Literature review

One of the central ideas behind this thesis is that the linear view of ‘the more self-control, the better’ is erroneous, and that excessive self-control (overcontrol) can be as problematic as insufficient self-control (undercontrol). While a Neurobiosocial Theory for Disorders of Overcontrol and a corresponding Radically Open Dialectical Behaviour Therapy (RO DBT), tailored to posited overcontrolled deficits and needs, have been proposed (Lynch, 2018), and the emerging evidence is promising, a lack of a valid and reliable instrument designed to identify maladaptive overcontrol among individuals is hindering the progress in the area. Therefore, the overarching aim of this thesis is to develop a new self-report instrument that will allow for accurate identification of overcontrolled individuals. The current chapter defines the construct of self-control and introduces the most prominent theories of self-control, including both those endorsing the linear view of self-control which does not recognise that excessive overcontrol can be problematic, and the quadratic view of self-control, which recognises that both under- and overcontrol can be problematic. Scales used to assess self-control in the context of the discussed theories are also reviewed. Special attention is paid to the Neurobiosocial Theory of Disorders of Overcontrol, which is a framework foundational to this thesis. Lastly, the rationale and research aims are provided.

2.1 The construct of self-control

Self-control has been considered a universal virtue for hundreds of years, with references to the high value of self-control dating back to ancient times. For instance, self-control was an important element of stoicism – a Hellenistic school of philosophy which originated in Athens around 300 BCE, strongly influenced by the teachings of philosophers such as Aristotle, Plato, and Socrates (Sellars, 2006). For stoics, exercising self-control (*enkrateia*) was a way of achieving discipline, emotional stability, and autonomy, and avoiding negative consequences of giving in to impulses and temptations (*akrateia* or *akrasia*; Kosman, 2020; Meyer, 1987). The high value placed on self-control was also expressed in traditional teachings of most mainstream religions, as means of inhibiting undesirable thoughts and behaviours and promoting behaviours that align with religious values and dogmas (Rounding et al., 2012). In Christian teachings, for instance, self-control was presented as a core virtue, a gift from

the Holy Spirit, necessary to overcome sinful temptations and live a moral life (Spiegel, 2020). Islamic teachings also stress the importance of restraint, discipline, patience, and the ability to regulate own behaviours, thoughts, and emotions (Rahman, 2018). The evolutionary importance of the ability to exert self-control has also been widely recognised, with researchers suggesting that self-control evolved to aid the survival and reproductive success of species (Green & Spikins, 2020). Data indicate that the cognitive capacity for self-control increased with brain volume as a response to intensified socioenvironmental pressures (MacLean et al., 2014). In humans, the ability to refrain from acting on self-serving impulses and temptations fosters cooperation and effective communication between individuals, thereby facilitating the formation of tribes and aiding interpersonal trust (Lynch, 2018; Righetti & Finkenauer, 2011). Further, self-control allows individuals to concentrate for long periods of time, delay gratification, and tolerate frustrations – skills crucial for effective prey hunting and early advancements in craftsmanship in early humans (Green & Spikins, 2020).

The ability to exert self-control remains important in modern contexts, with psychological research widely linking self-control failure to a range of negative life outcomes (see e.g., Cobb-Clark et al., 2022; Converse et al., 2018; Gottfredson & Hirschi, 1990; Moffitt et al., 2011; Tangney et al., 2004). The interest in self-control in modern social science research dates to 1930s (Baumeister et al., 2019), when Freud (1930) proposed that the ability to inhibit self-serving impulses and conform to the demands of the tribe was crucial to human adaptation (as cited in Tangney et al., 2004). However, the first comprehensive theoretical models of self-control only started to form towards the end of the 20th century (see Baumeister & Vohs, 2018). Over time, the research interest in the topic of self-control has grown substantially (Duckworth & Kern, 2011), with a quick search for “self-control” on Google Scholar returning approximately 2,610,000 results at the time of writing. Yet, despite a plethora of existing research on the construct, there is little agreement on what exactly the term “self-control” entails (Green & Spikins, 2020; Wennerhold & Friese, 2023). Authors commonly provide different definitions of the construct of self-control, largely depending on their area of expertise and view on personality development (Gillebaart, 2018). Commonly, the term ‘self-control’ is used to describe differing constructs (e.g.,

state versus trait self-control¹), used interchangeably with terms describing other constructs cross-sectionally closely related to self-control (e.g., impulsivity, disinhibition) without providing clear explanations of empirical differences between the constructs (Pilcher et al., 2023), or even used without being explicitly defined (Green & Spikins, 2020). This lack of an agreement across and within disciplines has been identified as one of the issues hindering further progress of research in the area (Pilcher et al., 2023). The sections that follow will consider different perspectives on self-control and clarify how the construct is conceptualised within the current thesis.

Of note, due to focus on dysfunction of personality, the thesis concentrates on trait self-control (i.e., as a relatively stable general tendency), rather than state self-control (i.e., how self-control increases and depletes depending on circumstances). A wider discussion on this is beyond the scope of this thesis (see e.g., Baumeister et al., 2019; De Ridder et al., 2018; Forestier et al., 2018; Wennerhold & Friese, 2023).

2.1.2 Self-control and self-regulation

Crucial to defining self-control is first establishing the boundaries between self-control and self-regulation. There is no universal agreement on whether and how self-control and self-regulation differ. The two terms have been used interchangeably by some researchers (Nigg, 2017) – for instance, the strength model of self-control (discussed in more detail later in this chapter) was referred to both as a model of self-control (e.g., Baumeister et al., 2007) and self-regulation (e.g., Baumeister et al., 2018). Others, however, have argued that they are distinct constructs, and have conceptualised self-control as a form of, or a component of, self-regulation.

For instance, Davisson and Hoyle (2017) described self-control a *conscious* self-regulatory strategy that allows specifically for inhibition of goal-irrelevant responses, initiation of goal-directed behaviour, as well as continuation of previously initiated goal-directed behaviour. In the authors view, the more general construct of self-regulation also includes processes that are automatic and cannot normally be consciously controlled, such as homeostatic processes. Similarly, Inzlicht and colleagues (2021) defined self-control as a component of self-regulation that serves to

¹ Trait self-control depends to one's dispositional level of self-control that is relatively stable over time, while state self-control is concerned with how the level of self-control within an individual varies over time depending on the circumstances (De Ridder et al., 2012).

resolve a real or anticipated conflict between two competing goals – for instance, a conflict between a desired outcome and a temptation, the pursuit of which may stand in the way of achieving that outcome.

Gillebaart (2018) attempted to systematically explain how self-control may function as a component of self-regulation. She defined self-regulation as a “system of standards, thoughts, processes, and actions that guide people’s behavior toward desired end states” that “provides the entire scaffolding for successful goal pursuit” (p. 3). She then proposed an ‘operational’ definition of self-control based on a Test-Operate-Test-Exit (TOTE) model (Powers, 1973) to explain how self-control fits within the broader construct of self-regulation. The TOTE model, aligned with the most prominent models of self-regulation, understands human behaviour to be goal-directed and controlled via feedback loops (Thürmer et al., 2020). The self-regulation feedback loop involves setting a goal, evaluating the discrepancies between the current state and the goal (*Test*), taking action to reduce these discrepancies in pursuit of the goal (*Operate*), and evaluating the effectiveness of the undertaken action (*Test*) – until the goal has been achieved (*Exit*; see Carver & Scheier, 1981, 1982; Davisson & Hoyle, 2017). Gillebaart proposed that self-control functions as the *Operate* component of the self-regulation cycle, aiming to reduce the discrepancies between the state and the goal through adjusting one’s responses and behaviour. In this thesis, in line with Gillebaart’s view, self-control is understood to be an operational component of self-regulation that serves to alter one’s responses and behaviours in a way that, in the actor’s view, allows for a successful pursuit of a desired goal.

2.1.2 The components of self-control.

Inhibition and initiation.

Despite self-control being a complex construct, available studies overwhelmingly consider self-control to be a primarily inhibitory mechanism (De Ridder et al., 2011). Inhibitory control has been defined as the ability to “control one’s attention, behavior, thoughts, and/or emotions to override a strong internal predisposition or external lure, and instead do what’s more appropriate or needed” (Diamond, 2013, p. 137) in pursuit of desired outcomes or goals. The ability to inhibit has long been considered a cognitive function crucial to adaptive functioning, due to inhibitory failure being associated with negative life outcomes, psychological and neurodevelopmental disorders, and criminal

activity (e.g., Billieux et al., 2010; Fino et al., 2014; Jasinska et al., 2012; Lipszyc & Schachar, 2010; Smith et al., 2014; Verona et al., 2012). However, De Ridder and colleagues (2011) highlighted that the ability to successfully pursue desired goals and outcomes (e.g., losing weight) relies not only on the ability to inhibit goal-irrelevant responses (e.g., suppress the impulse to reach out for chocolate), but also the ability to stimulate goal-directed behaviour (e.g., make an effort to exercise regularly and prepare healthy meals). The researchers argued that despite the existence of frameworks that explicitly recognise inhibition and initiation as two distinct types of self-control, research in the area disproportionately focuses on inhibition while overlooking the importance of initiatory control. To support their claim, they conducted a factor analysis to establish whether it is possible to empirically differentiate between inhibitory and initiatory self-control. Indeed, the results of the study supported a two-factor model, with one factor representing inhibitory behaviours, and one factor representing initiatory behaviours. In line with the empirical evidence provided by De Ridder and colleagues' (2011), the current thesis recognises that adaptive self-control is not simply equivalent to successful inhibitory control and considers initiation as one of the core components of the construct of self-control.

Effort.

Self-control is also commonly discussed in terms of effort. Effortful control refers to the ability to inhibit responses and impulses (or initiate actions) *voluntarily* (Eggum-Wilkens et al., 2016). The most prominent model of self-control – i.e., the strength model (Baumeister et al., 1994, 1998) – posits that exerting self-control always requires effort (Baumeister et al., 2007), thereby implying that the decision to exert self-control is conscious (Gillebaart & De Ridder, 2015). However, it is not universally agreed that all self-control responses require effort. Gillebaart and De Ridder (2015) argued that individuals characterised by high levels of self-control may be more likely to pursue goals via use of habitual strategies and automated behaviours that require relatively low effort – such as pro-active temptation avoidance (e.g., avoiding the candy alley in a supermarket while trying to limit sugar intake). They argued that while such behaviour may require effort at first, when repeated, would quickly become habitual and automatised in people with good self-control, and thereby no longer requiring effort. The researchers also posited that people with higher self-control may more efficiently down-regulate the conflict between wanting to pursue a small instant

reward over a larger distant reward. This is by attributing hedonic properties to the object aligned with the distant goal to the same extent as to the object of the immediate temptation, resulting in equal hedonic activation of both options. Effectively, individuals who can apply self-control successfully likely do not experience temptations as intensely, resulting in little to no effort required to avoid the temptations. An alternative explanation could be that people high in self-control present with lower reward sensitivity when compared to those with lower self-control, as proposed by Lynch (2018). Little sensitivity to reward could mean that highly controlled people would not experience the response conflict in the first place, and therefore no effortful response would be required. To allow for the possibility that effort may not always be involved in self-control behaviours, the definition of self-control within this thesis includes all attempts at altering behaviours and responses, whether effortful or effortless, made in pursuit of a desired outcomes or goals.

Affect regulation.

There has been some discussion regarding whether affect regulation – encompassing emotion regulation, mood regulation, and regulation of stress responses (Gross 2008, 2014) – should be considered as a component of self-control. Green & Spikins (2020) noted that self-control is usually discussed in a ‘cold’ way – referring to cold executive functions, such as planning, organising, inhibiting goal-irrelevant behaviour, or staying focused. However, Gross (1998; 2014) posited that the kind, valence, and intensity of emotions depends on the individual’s interpretation of a situation in relation to their currently active goal. He argued that emotions not only make people ‘feel’ a certain way, but also act a certain way, and in that sense, they may help or interfere with achieving a desired outcome. For example, in emotionally charged situations, hot executive functions may guide individuals to make decisions and control emotionally driven impulses that directly support their active goals – such as choosing whether to argue or walk away during a heated conflict. Thereby, it is reasonable to hypothesize that individuals may attempt to alter their affective response in a way that they believe will help them achieve a desired outcome (e.g., openly express sadness to make someone behave in a desired way, or push-down the feelings of distress resulting from working long hours when an important deadline is approaching). Moreover, Green and Spikins (2020) argued that inhibition cannot be considered without considering *deliberate* regulation of affect, due to neuroanatomical overlap between the two (see

e.g., Etkin et al., 2015), advocating for the umbrella term of self-control to include *conscious* affect regulation. It is less clear whether automatic control of emotion outside of consciousness should also be considered part of the construct of self-control. Though, in line with Gillebaart and De Ridder's (2015) position that self-control responses could become habitual under certain conditions, it can be hypothesized that affect regulation responses could also become automatic in a similar way. Considering the above discussion, in this thesis, emotion regulation (conscious or automatic) is conceptualised as a (secondary) component self-control and defined as an attempt at altering an affective state to answer the demands of the environment that the actor perceives as to be beneficial in pursuit of desired outcomes or goals.

2.1.8 The current thesis' definition of self-control.

To conclude the previous sections, in this thesis, the term 'self-control' denotes a trait-like, multi-faceted construct, and it is considered an operational component of self-regulation. Throughout, it is used to describe any response or strategy – whether inhibitory or initiatory, and whether requiring any effort – applied by an individual in pursuit of a desired outcome or goal. This includes responses aimed at altering and regulating behaviours, cognitions, thoughts, impulses, and emotions.

2.2 Linear view of self-control

Despite widely differing definitions of the construct of self-control, one thing that much of the research on self-control has in common is the belief that *good* self-control equals *high* self-control, and *bad* self-control equals *low* self-control. The ability to exert self-control has been empirically linked with a range of positive life outcomes within modern psychological literature (Bohane et al., 2017; De Ridder et al., 2012; Lynch, 2018). For instance, it was linked to better physical and mental health (Tangney et al., 2004), greater life satisfaction and happiness (De Ridder et al., 2011; Hofmann et al., 2014), academic and workplace success and higher self-esteem (Tangney et al., 2004), better job prospects and greater financial well-being (Daly et al., 2015; Moffit et al., 2011) higher quality interpersonal relationships (Vohs et al., 2011), better coping skills (Gailliot et al., 2007), and more constructive responding in challenging interpersonal situations (Finkel & Campbell, 2001).

Conversely, too little control over impulses, behaviours, and emotions – in this thesis, referred to as *undercontrol* – has been widely linked to negative life outcomes, such as substance misuse, unsafe sex, teen pregnancy, and drink-driving (Moffitt et al., 2011), impulse-buying (Baumeister, 2002), emotional eating and binge-eating (Elfhag & Morey, 2008; Pearson et al., 2018), and financial debt (Gathergood, 2012). Those low in self-control have also been shown to be more likely to, in extreme cases, experience a range of complex psychopathological outcomes and engage in violent, aggressive, and criminal behaviour (Caspi et al., 1996; Daly et al., 2015; De Ridder et al., 2012; Gottfredson & Hirschi, 1990; Krueger et al., 1996; Moffitt et al., 2011; Vaughn et al., 2007, White et al., 1994).

In line with the above findings, the relationship between self-control and personal well-being is widely believed to be linear – i.e., the more capacity for self-control the individual has, the less likely they are to experience adverse life outcomes, have ill mental health, or engage in criminal activity (Hamilton, 2021). This linear view of self-control was endorsed by numerous self-control models proposed over the years, with the two most influential being Baumeister and colleagues' (1994; 1998) strength model of self-control and Gottfredson and Hirschi's (1990) General Theory of Crime. These two theories were arguably the most influential in popularising the linear view of self-control.

The belief 'the more self-control, the better' stems largely from studies that link low self-control to negative life outcomes that endorsed views rooted in the two theories and utilised measures developed based on their assumptions. Due to the profound impact of the strength model and the General Theory of Crime on the current understanding of self-control, the paragraphs that follow in the current section concentrate on critically discussing these two theories. The key assumptions of the theories are outlined, and assessment measures commonly used in studies that endorse their understanding of self-control are considered.

2.2.1 The strength model of self-control.

The strength model of self-control was proposed by Baumeister and colleagues (1994; 1998; see also Muraven & Baumeister, 2000; Muraven et al., 1998). Based on a review of the available literature, the researchers proposed that self-control depended on a

limited energy resource and was vulnerable to depletion. They compared self-control to the workings of a skeletal muscle. That is, when a skeletal muscle is used, it becomes fatigued and requires time to recover before it can function properly again. According to the strength model, the same happens with self-control – when self-control is exerted, the capacity for self-control becomes diminished, temporarily reducing one's capacity for self-control. This state of reduced self-control capacity is known as ego depletion (Baumeister et al., 2007; Hagger et al., 2010). Studies show that the state of ego depletion can be offset through rest and relaxation, or by glucose supplementation (Galliot et al., 2007; Galliot & Baumeister, 2007; Tyler & Burns, 2008). Further, the strength model proposes that just as skeletal muscles can be strengthened with regular training, the capacity for self-control can also be increased through repeated exercise (Muraven, 2010; Muraven et al., 1999) – and presents this as a potential explanation for individual differences in trait self-control across individuals.

The researchers who endorse the strength model consider the capacity to exert self-control to be “one of the most powerful and beneficial adaptations of the human psyche” (Tangney et al., 2004, p. 272), and support the view that insufficient trait self-control underlies a majority of personal and social problems individuals face (Baumeister & Vohs, 2004). This is in accordance with the linear view of the construct of self-control, with Baumeister and colleagues (1994; 2009) questioning the concept of excessive trait self-control, and only acknowledging the transient limitations of the ego depletion state. Notably, most of the research on the strength model focuses on the inhibitory function of self-control, defining self-control as the capacity to overcome impulses and temptation through consciously modifying one's behaviour (Hagger, 2010). This is despite Baumeister and colleagues (1998) recognising that self-control also functions to initiate desired behaviours (De Ridder et al., 2011).

The strength model studies typically concentrate on testing hypotheses regarding the mechanisms behind ego-depletion, as well as the ways to offset the effect and increase capacity for self-control – with self-control failure understood as the inability to apply self-control when it would be advantageous (Hagger et al., 2010). The studies typically utilise behavioural tasks (Hagger et al., 2010) – such as executive function tasks and delay of gratification tasks (Duckworth & Kern, 2010). Executive function tasks are designed to assess goal-directed cognitive processes that influence

task performance (e.g., working memory, attention, task-switching, and inhibition; Duckworth & Kern, 2011). Executive function tasks famously utilised in assessing self-control capacities include the go/no-go task (see Donders, 1969; LaBerge & Samuels, 1974), the Stroop task (Stroop, 1935), the stop-signal task (Logan & Cowan, 1984), and the flanker task (Eriksen & Eriksen, 1974). Better performance on the tasks indicates better impulse control, translating to better self-control. In delay of gratification tasks, in turn, participants are required to make choices between a smaller immediate reward and greater delayed reward (Duckworth & Kern, 2011), with greater capacity for delaying gratification indicating better self-control. These include delay discounting tasks (see Matta et al., 2012) and the renowned marshmallow test (Mischel et al., 1988).

A meta-analysis of studies testing the assumptions of the strength model of self-control in relation to task performance indicated a high level of convergence across studies' results, providing support for the ego-depletion hypotheses (Hagger et al., 2010). However, a limitation of using experimental tasks to measure self-control abilities is that each task is only designed to measure a single, narrow domain of self-control (e.g., impulse control). As a result, the tasks do not always substantially correlate, putting into question whether they measure the same construct (Duckworth & Kern, 2011; Wennerhold & Friese, 2020). Further, even if the tasks do measure different aspects of the construct of self-control, they certainly do not capture the full extent and complexity of the construct. As such, they are not suitable for testing the hypotheses regarding the links between self-control and other personality characteristics and life outcomes that the strength model supporters endorse. For this purpose, self-report scales are a viable alternative, as they allow to examine trait self-control more comprehensively and in a more resource-efficient way².

The most prominent self-report scale used to measure the strength of self-control in accordance with the linear view is Tangney and colleagues' (2004) Self-

² The researcher acknowledges that informant-report measures can outperform self-report measures and that combining both self- and informant-report measures is recommended when feasible to enhance the reliability of personality assessment (Duckworth & Kern, 2011; Möttus et al, 2024; Paulhus & Vazire, 2007). However, particularly with highly overcontrolled individuals, reaching suitable informants may be challenging for researchers and clinicians, due their hypothesised lack of social connectedness, emotional loneliness, low openness, and the fear of vulnerability (Lynch, 2018). As such, informant-report measures are not considered further within this thesis.

Control Scale (SCS) – or its short version, the Brief Self-Control Scale (BSCS; see section 7.2.3 for a more detailed description of the scale, including scoring and psychometric properties). It conceptualises self-control as a unidimensional construct, with people scoring higher on the scale understood to have greater trait self-control strength (Englert, 2021). The scale is commonly used to explore the relationship between self-control and other psychological constructs and personality traits, as well as the links between individuals’ capacity for self-control and life outcomes. The findings of studies utilising the SCS show overwhelming support for the claims that low self-control underlies a variety of negative outcomes, while ‘high’ self-control is a protective factor – with the original study alone reporting high self-control to be linked to better grades, more satisfactory relationships, better interpersonal skills, higher self-esteem, more secure attachment, more optimal emotional responding, and less psychopathology. A meta-analysis by De Ridder and colleagues (2012) indicated that the SCS was able to predict both inhibition of undesired and instigation of desired behaviours (despite including more items relating to the former), as well as allowed for a more comprehensive exploration of the relationship between the impact that self-control has on one’s behaviours and the associated life outcomes when compared to other self-control scales.

Nevertheless, there are several limitations associated with using the SCS to measure self-control. Firstly, it conceptualises self-control as a unidimensional construct. This view is rather outdated, with more recent research emphasising the multidimensional nature of self-control (Hamilton, 2021). As such, the SCS may not capture the breadth and complexity of the construct of self-control. Secondly, while the SCS claims to measure the spectrum from low to high self-control, the items on the scale focus on adaptive behaviours associated with the ability to exert self-control (“*I am good at resisting temptation.*”) and maladaptive behaviours characteristic to those with insufficient self-control (e.g., “*I often act without thinking through all the alternatives.*”). Despite the concept of overcontrol being present in psychological literature at the time, items that would capture the maladaptive behaviours theorised to be associated with high self-control are lacking. It is therefore likely that rather than measuring the spectrum from low to high self-control, the scale measures the spectrum from low to good (i.e., adaptive) self-control. This is interesting given that in their original paper, Tangney and colleagues (2004) conducted a set of analysis aiming to

examine whether the excess of self-control can also be problematic. The researchers failed to find a curvilinear relationship between self-control and psychological adjustment, and thereby concluded that there was no evidence for the concept of too much self-control. An alternative explanation is that the curvilinear relationship was not found due to the scale lacking in items written in a way that would allow to specifically assess maladaptive overcontrol and would distinguish it from undercontrol. Still, the prominence of the strength model of self-control and the wide use of the SCS have largely contributed to the popularity of the linear view of self-control, thereby lessening the interest in the possible negative effects of excessive self-control.

2.2.2 The General Theory of Crime.

The General Theory of Crime by Gottfredson and Hirschi (1990) is another prominent theory of self-control that endorses a linear view of the construct. It was introduced in the field of criminology as an attempt to explain why individuals engage in offending behaviour and ‘crime analogous’ behaviour, such as substance use, gambling, and other high-risk, reckless actions and activities (Hirschi & Gottfredson, 1988; Pratt & Cullen, 2000). The theory presents the low trait self-control as the sole basis for the development and maintenance of criminal and other deviant behaviour (Buker, 2011), based on the assumption that committing crime and engaging in delinquent behaviour is a simple way to achieve immediate gratification (Pratt & Cullen, 2000). It defines low self-control as a tendency to pursue short-term gratification, and simultaneously discount potential negative long-term effects of one’s actions (Venables et al., 2018).

Gottfredson and Hirschi (1990) described those characterised by low trait self-control as impulsive, rash, reckless, and insensitive individuals that cannot tolerate frustrations and favour simple and physical tasks and activities over complex and intellectual ones. They theorised that there are six distinct components of low trait self-control – impulsivity, risk seeking, temperament, self-centeredness, preference for simple tasks, and preference for physical activities – that work together and remain stable over the life course. Despite having distinguished several separate components of low self-control, due to their interdependence, the theorists conceptualised the trait as a unidimensional construct.

To measure low trait self-control in line with Gottfredson and Hirschi's (1990) conceptualisation, the Low Self Control Scale (LSCS) was developed by Grasmick and colleagues (1993; see section 7.2.3 for a more detailed description of the scale, including scoring and psychometric properties). The scale was designed to contain items that would reflect behaviours and characteristics of individuals associated with each of the six components, but still conceptualised to be a unidimensional tool. This is because the scale developers failed to find convincing evidence of the multidimensionality of the scale, with the components apparently representing a single construct, seemingly validating the claims of the underlying theory (Pechorro et al., 2023). It has since been the most common self-report assessment tool used to test the assumptions of the General Theory of Crime, with cross-sectional and longitudinal studies largely substantiating the link between low self-control and offending and deviant behaviour (Pechorro et al., 2023; Pratt & Cullen, 2000; Vazsonyi et al., 2017).

Notwithstanding the supporting body of evidence, both the theory and the scale have been widely criticised. Gottfredson and Hirschi's (1990) proposition that a single personality trait of low self-control explains criminal behaviour has been criticised for being parsimonious and reductionist, and the conceptualisation of self-control as a unidimensional construct critiqued for being outdated (Hamilton, 2021; Marcus, 2004). Hamilton (2021) argued that the linear view likely to be erroneous in the light of available theoretical and empirical literature on overcontrol, and the theory's assumption that self-control can be viewed as a unidimensional construct is not up to date with contemporary theoretical perspectives. Indeed, some studies have failed to replicate the unidimensional structure of the LSCS (e.g., Piquero et al., 2000), indicating that the view on self-control on which the scale was based may be erroneous (Venables et al., 2018). Further, psychometric studies aiming to establish the validity and reliability of the LSCS have provided inconsistent results (Pechorro et al., 2023). While some studies indicated good validity and reliability cross-culturally and independently of characteristics such as sex and age (Pechorro et al., 2023; Vazsonyi & Belliston, 2007), other studies criticised the scale for being gender-biased (Gibson et al., 2010), affected by race and language (Ward et al., 2018), and questioned its dimensionality and construct validity (Higgins, 2007; Marcus, 2004). Moreover, while some of the six dimensions of low self-control proposed by Gottfredson and Hirschi are widely accepted to be crucial to the construct of self-control in the forensic context,

others (particularly the preference for simple task and physical activity) have been suggested to be less relevant (Pechorro et al., 2023). Lastly, some of the proposed dimensions have also been argued to likely be consequences of low self-control rather than the core components of the construct, and the scale has been argued not to reflect the breadth of the construct that it claims to measure (Hoyle & Davisson, 2017).

Despite the criticisms of the General Theory of Crime and the inconsistent evidence as to the validity and reliability of the LSCS, the popularity of the theory only increased over the years, causing the linear view of self-control to become dominant also in forensic contexts (Hamilton, 2021). Consequently, Grasmick and colleagues' (1993) scale based on the theory remains a common choice of a self-control assessment in empirical studies (Pratt & Cullen, 2000; Vazsonyi et al., 2017) and is even considered a gold-standard scale in the field of criminology due to the frequency of its use (Pechorro et al., 2023).

2.4 Quadratic view of self-control

With the linear view of self-control prevailing, unsurprisingly, the literature on adverse effects of self-control has focused almost exclusively on undercontrol. This is despite the notion of problematic overcontrol having been present in psychological literature for several decades – and proposed for the first time well over a decade before the General Theory of Crime or the strength model of self-control were published. The paragraphs that follow discuss two quadratic perspectives on overcontrol – Megargee's (1966) idea of overcontrolled violent offending and Block and Block's (1980) idea of the Resilient-Overcontrolled-Undercontrolled (RUO) typology (including considerations of the concepts of ego-control and ego-resilience).

2.4.1 Megargee's idea of under- and overcontrolled offending.

The idea that excessive self-control can lead to adverse life outcomes was first proposed in a forensic context by Megargee (1966). He theorised that there were two types of individuals who commit violent offences – aggressive undercontrolled and chronically overcontrolled. He portrayed the aggressive undercontrolled individuals as impulsive, disinhibited, violent, easily frustrated, and highly responsive to provocation – a description closely resembling that of individuals characterised by low self-control in Gottfredson and Hirschi's (1990) theory. The chronically overcontrolled individuals

were described by Megargee as seemingly prosocial, quietly suffering individuals with superior control over their behaviours and emotions. In his view, it was the hidden suffering and pervasive masking of their true feelings that ultimately led the overcontrolled to commit offences in a sudden outburst of aggression associated with self-control depletion. Further, Megargee theorised not only about personality characteristics of under- and overcontrolled individuals who commit offences, but also about types of offences they are likely to commit. He suggested that the aggressive undercontrolled individuals, because of their impulsive nature, would have lengthy forensic histories and engage in both violent and non-violent offences. In contrast, he expected the chronically overcontrolled individuals to commit one-off offences that were always violent in nature and often catastrophic in consequences.

Megargee's (1966) theorising initially gained quite some traction among researchers, with cross-sectional studies indicating that a substantial proportion of people across forensic settings may be overcontrolled, reporting prevalence rates of 16-52 percent (D'Silva & Duggan, 2010; Hamilton, 2021; Hempel et al., 2018; see e.g., D'Silva & Duggan, 2010). However, while Megargee in his work theorised about clinical presentations of under- and overcontrolled individuals who commit violent offences, he did not provide a solid theoretical framework explaining the mechanisms of under- and overcontrolled offending that could be empirically tested (Hamilton, 2021). Further, the clinical descriptors provided by Megargee were highly specific, and therefore difficult to replicate and easily and widely challenged. For instance, researchers identified overcontrolled tendencies among individuals who committed non-violent offences, contradicting Megargee's claims that overcontrol is a violent offending phenomenon, and many studies have failed to provide conclusive evidence regarding mechanisms that would underpin a specific overcontrolled offending pathway (Hamilton, 2021; see e.g., McGurk & McGurk, 1979; Truscott, 1990). Additionally, the validity and reliability of the Overcontrolled Hostility Scale – a scale that Megargee and colleagues (1967) derived from the Minnesota Multiphasic Personality Inventory (Hathaway & McKinley, 1967) to assess overcontrol – were widely questioned (e.g., Hoppe & Singer, 1976; Hutton et al., 1992; McCreary & Padilla, 1977). Further, the scale was often used incorrectly, with low scores assumed to measure undercontrol, leading to confounding results (Hamilton, 2021; e.g., Du Toit & Duckitt, 1990; Hershorn & Rosenbaum, 1991). Due to the prominent criticisms of

Megargee's theorising as well as issues around the properties of the Overcontrolled Hostility Scale, neither the theory or the measure are considered further within this thesis. For a recent test of Megargee's hypotheses and a discussion on the issues around his theorising and how they led to the interest in overcontrolled offending gradually fading, see Hamilton (2021).

2.4.2 Ego-control and ego-resilience.

In non-forensic psychology literature, the concept of overcontrol was first proposed in 1971 by Block. In the study, utilising longitudinal archive data (Pulkkinen, 1996) and the Q sort method (Block, 1961), Block identified five distinct personality types in a sample of 84 men, three of which were found to remain stable from adolescence to adulthood (Robins et al., 1996):

- (1) Ego Resilients – well-adjusted, interpersonally effective individuals,
- (2) Unsettled Undercontrollers – impulsive, antisocial individuals,
- (3) Vulnerable Overcontrollers – rigidly controlled individuals characterised by low adaptability.

Corresponding types were identified in a study by Robins and colleagues (1996), and labelled Resilients, Undercontrollers, and Overcontrollers (RUO types) in 300 adolescent boys. The Resilients were shown to be intelligent and academically successful, prosocial, and unlikely to cause trouble or suffer from ill mental health. The Overcontrollers shared some characteristics with the Resilients – for example, both groups were well-behaved and unlikely to engage in 'delinquent' behaviours. However, the Overcontrollers, despite being as intelligent as the Resilients, were not as successful in school, less likely to engage socially, and more likely to suffer from internalising mental health problems, such as anxiety. The Undercontrollers, in turn, were found to score significantly lower on the IQ scale compared to the other two groups, exhibited the highest levels of dysfunction, were prone to a range of academic, social, behavioural, and emotional issues, and both internalising and externalising symptomatology. The RUO types (although with some variation in the number of types and the characteristics within types) have since been commonly identified in both men and women, across age groups, and in both general and clinical populations –

highlighting the importance of recognising both under- and overcontrolled problems when designing and administering psychological interventions (Bohane et al., 2017).

The RUO types were shown to differ on the dynamic dimensions of ego-functioning identified by Block and Block (1980) – the ego-control and ego-resilience dimensions (Robins et al., 1996). Ego-control refers to one's typical response to behavioural and attentive impulses – their inhibition or expression (Letzring et al., 2005; Oshio et al., 2018). In accordance with the original RUO study (Block, 1971), Block (1993) and Block and Block (1980) conceptualised ego-control as a dimension from under- to overcontrol (Letzring et al., 2005). Block (1971) described undercontrol as a behavioural tendency towards impulsivity, expressivity, spontaneity, openness to experience, low levels of conformity, high tolerance of ambiguity and change, frequent but often short-lived enthusiasm about a variety of activities, and global rather than detailed-focused processing of information. Overcontrol, in turn, he linked to high constraint, rigidity, behavioural and emotional inhibition, detailed-focused processing, low openness, the preference for delaying gratification, high perseverance, high conformity, and the tendency to carefully plan and avoid ambiguity (Block, 1971). Ego-resilience, in turn, refers to the individual's dynamic capacity to temporarily modify their level of ego-control depending on the environmental and situational context (Letzring et al., 2005; Oshio et al., 2018). Block posited that both highly undercontrolled and high overcontrolled individuals would be set in their ways, and thereby low in the ability to adapt to changing circumstances – i.e., low in ego-resilience. Meanwhile, individuals high in ego-resilience were expected to fall further away from either of the extreme ends of the ego-control spectrum. This theorising was supported by the findings of the study by Robins and colleagues (1996), showing that highly resilient adolescent boys exhibited intermediate levels of ego-control, while both highly undercontrolled and highly overcontrolled boys scored low on ego-resilience. The study also linked low levels of ego-resilience to negative life outcomes, with 'ego-brittle' boys exhibiting considerable levels of dysfunction and maladjustment. Other studies also support the beneficial nature of high ego-resilience, linking it to, for instance, more adaptive expression of emotions, better adaptive functioning, better emotional well-being, higher empathy, better interpersonal relationships and overall social competence, more effective coping strategies, less psychopathology (especially internalising), and less behavioural issues (Block &

Block, 1980; Causadias et al., 2012; Cumberland-Li et al., 2004; Letzring et al., 2005; Milioni et al., 2015; Ong & Bergeman, 2004; Oshio et al., 2018; Philippe et al., 2011).

However, the interest in studying ego in the context of personality gradually diminished along with the growth in popularity of alternative, psychometrically-derived theoretical models of personality, such as the Five Factor Model (FFM; Costa & McCrae, 1992; Goldberg, 1990; McCrae & John, 1992; see Waaktaar & Torgersen, 2010). While several studies have considered the RUO types in the context of the FFM, the results have been largely inconsistent regarding the number and content of personality prototypes, with researchers most commonly identifying either three, four, or five personality types, and variable configurations of factors across the types. The Resilient type has generally been indicated to present with positive characteristics (i.e., those widely linked to high adaptability and low psychopathology) with regards to all five factors – low neuroticism, high extroversion, high conscientiousness, high openness to experience, and high agreeableness (Robins et al., 1996; Strus et al., 2021). There is, however, less clarity as to under- and overcontrolled trait configurations. While both constructs are consistently linked to more maladaptive patterns of the FFM traits, various studies provide diverse results as to the configuration of the traits within the types (see Strus et al., 2021). Strus and colleagues (2021) attribute these inconsistent results to the exploratory nature of the RUO personality typology studies. Additionally, the variability in the personality measures utilised across the studies also likely contributes to the lack of consensus on the number of types and configuration of traits within types.

This diminished interest in studying the constructs of ego-resilience and ego-control (especially in the context of overcontrol) are also partially a result of issues with measuring the constructs. Firstly, the measurement of ego-resilience is considered. Early studies on the theory used a descriptive, judge-rated, and time-consuming Q-sort method (Block, 1978) of group assignment (Alessandri et al., 2007). However, with the rise in popularity of self-report questionnaires, Block and Kremen (1996) published and examined a Likert-type, unidimensional scale to assess individual differences in ego-resilience – known as the Ego-Resilience Scale-89 (ER-89) – designed for use in non-clinical populations (Alessandri et al., 2007; see section 7.2.3 for a more detailed description of the scale, including scoring and psychometric

properties). The items for the ER89 came from a variety of sources, many of which are currently untraceable (Alessandri et al., 2007). Initial validity evidence that Block and Kremen reported for the ER89 was based on the longitudinal study on ego development by Block and Block (1980). The scale was then further validated by Letzring and colleagues (2005), yielding satisfactory psychometric properties. Since then, the ER89 scale has been widely used to measure ego-resilience beyond the general population. However, later studies that examined the dimensionality of the scale yielded mixed results, with some failing to find evidence that it is unidimensional (e.g., Alessandri et al., 2007; Fonzi & Menesini, 2005). This is likely due to Block and Kremen erroneously basing their assumption of unidimensionality of the ER89 on the value of the alpha coefficient (Alessandri et al., 2007; see Hoekstra et al., 2019). This is problematic, as both under- and overestimation of the number of factors has been suggested to negatively affect the reliability of scales (Hayton et al., 2004). While some revised, multidimensional versions of the ER89 were developed, most of the available research utilises the original version – likely due to the lack of a theoretical basis that would explain the multidimensional nature of ego-resilience indicated by the revised models.

Letzring and colleagues (2005) noted that aside from the ER89, Block also attempted to develop a measure of individual differences in ego-control – later named the Ego-Undercontrol (EUC) scale. The scale was first published and examined in the study by Letzring and colleagues, and it is unclear when exactly the EUC scale was developed and why it was not published alongside the ER89. Unfortunately, Letzring and colleagues' examination of the scale indicated the scale's psychometric properties to be less than satisfactory. The lack of an instrument to measure the spectrum from under- to overcontrol resulted in studies employing various methods to categorise people as under- and overcontrolled (or resilient). The variety of methods used – and the fact that other self-control scales, as previously discussed, fail to account for overcontrol tendencies – likely contributed to the RUO studies yielding mixed results, leading to the interest in further researching the typology weakening over the years. More recently, Isaksson, Ghaderi, Wolf-Arehult, and Ramklint (2021) recognised the lack of a measure that would allow to assess overcontrol as a major problem. They decided to re-examine the EUC in a Swedish sample, and once again, the psychometric properties turned out to be less-than-satisfactory. They then went on to develop a short

version of the EUC – the EUC-13 – to improve the psychometric properties (see section 7.2.3 for a more detailed description of the scale, including scoring and psychometric properties). The new version of the scale generated more promising results in terms of the model fit and internal consistency, with only one of the three dimensions failing to meet the internal consistency reliability cut-off. However, it has not yet been used in other studies and populations.

Lastly, when considering limitations pertaining to Block's theorising, it is necessary to also note that associated studies emphasized the importance of environmental influences in shaping later personality functioning in the context of ego-control and ego-resilience – and especially the early developmental experiences. For instance, in girls, parenting characterised by strict behavioural control, discouragement of emotional expression, and focus on performance and outcomes was linked to ego-overcontrol later in life (Kremen & Block, 1998). In boys, having hostile and competitive mothers and impatient, ineffective, uninvolved fathers embarrassed about their sons was linked to later undercontrol (Kremen & Block, 1998). In mixed-sex samples, supportive parenting was linked to higher ego-resilience (Swanson et al., 2011), while insecure attachment in parent-child relationships was linked to lower ego-resilience (Caldwell & Shaver, 2012). Interestingly, while Block (2002) did recognise the possible role of genetic predispositions in the development of personality, he did not comprehensively address the topic in any of his works. In fact, he argued that separating genetic and environmental influences and exploring the interplay between the two in the context of personality was not crucially important. The downplaying of the role of biotemperament and genetic predispositions can be interpreted as another limitation of the framework, as it is commonly recognised that nature and nurture are integral in the context of personality development and dysfunction (e.g., Kandler & Zapko-Willmes, 2017; Plomin, 1994).

2.5 The Neurobiosocial Theory for Disorders of Overcontrol

Despite previous literature clearly supporting the notion that problematic overcontrol is present and prevalent among different populations, it has not received nearly as much attention as problematic undercontrol in both research and clinical contexts. The previous paragraphs have highlighted the likely reasons for the neglect of the construct overcontrol: the importance that the society places on self-control capacities, the

popularity of models that endorse a linear view of self-control, the lack of a solid theoretical framework that would comprehensively explain overcontrol, the mixed results of typological studies attempting to differentiate overcontrol from undercontrol, and the lack of a valid and reliable way to measure the construct.

Markedly, the diminishing interest in further researching and treating maladaptive overcontrol despite the evidence of its existence carries significant negative consequences – it means that individuals who experience difficulties related to overcontrol may not receive appropriate psychological help. Most prominent interventions that are currently widely offered to clients, such as Dialectical Behaviour Therapy, Cognitive Behaviour Therapy, Acceptance and Commitment Therapy, and Compassion Focused Therapy, put emphasis on enhancing self-control and increasing emotion regulation abilities (Heath et al., 2021; Luoma et al., 2018). However, researchers suggest that such interventions can be counterproductive or even detrimental to individuals who already pervasively inhibit emotions and apply self-control excessively (Day et al., 2005; Hamilton, 2021; Luoma et al., 2018; Lynch, 2018).

This presents a clear need for a new, verifiable, and comprehensive theoretical framework that would not only recognise, but also systematically define maladaptive overcontrol and explain the mechanisms behind the development and maintenance of problematic overcontrolled behaviours, relating to both nature and nurture. Such a framework, if thoroughly empirically tested, could provide a better understanding of maladaptive overcontrol, and guide the development of effective interventions tailored to the specific needs of overcontrolled individuals. And indeed, in 2018, a novel theory of maladaptive overcontrol – a Neurobiosocial Theory for Disorders of Overcontrol – was proposed by Thomas Lynch, developed based on more than 25 years of translational research.

2.5.1 An overview of Lynch’s theoretical framework.

Lynch’s (2018) Neurobiosocial Theory for Disorders of Overcontrol is based on the premise that adverse life outcomes, including complex psychopathology, stem from personality dysfunction. The theory defines personality dysfunction as *“habitual perceptual and regulatory biases that are either overcontrolled or undercontrolled in*

nature” (p. 69). As such, overcontrol and undercontrol are understood as superordinate personality types. Lynch posited that every person is biologically predisposed towards under- or overcontrol. Importantly, he postulated that undercontrolled nor overcontrolled personality style is inherently problematic. This is because most people are likely to be able to flexibly adapt their level of self-control depending on the demands of the situation. He suggested, however, that bidirectional interactions between certain temperamental dispositions, environmental factors, and coping mechanisms may make both under- and overcontrol maladaptive – i.e., lead to severe socio-relational and occupational difficulties and complex psychopathology (Diagnostic and Statistical Manual of Mental Disorders [DSM-5], American Psychiatric Association, 2013).

With the links between undercontrol and adverse life outcomes having been widely researched over the past decades, Lynch (2018) focused on explaining the behavioural manifestations and mechanisms behind maladaptive overcontrol. He defined maladaptive overcontrol by proposing four core deficits (p. 8):

*“1. **Low receptivity and openness**, manifested by low openness to novel, unexpected, or disconfirming feedback; avoidance of uncertainty or unplanned risks; suspiciousness; hypervigilance regarding potential threats; and marked tendencies to discount or dismiss critical feedback*

*2. **Low flexible control**, manifested by compulsive needs for structure and order; hyper-perfectionism; high social obligation and dutifulness; compulsive rehearsal, premeditation, and planning; compulsive fixing and approach coping; rigid rule-governed behaviour; and high moral certitude (the conviction that there is only one “right” way of doing something)*

*3. **Pervasive inhibited emotional expression and low emotional awareness**, manifested by context-inappropriate inhibition of emotional expression (for example, presentation of a flat face in response to a compliment) or by insincere or incongruent expressions of emotion (for example, a smile in response to distress, or a show of*

concern when no concern is actually felt); consistent underreporting of distress; and low awareness of bodily sensations

*4. **Low social connectedness and intimacy with others**, manifested by aloof and distant relationships; a feeling of being different from other people; frequent social comparisons; high envy and bitterness; and reduced empathy.”*

The Neurobiosocial Theory for Disorders of Overcontrol (Lynch, 2018) aims to systematically explain how the above deficits and, effectively, maladaptive overcontrol, may develop and be maintained among individuals. The theory is comprised two separate yet interrelated models: (1) the Neuroregulatory Model of Socioemotional Functioning and (2) the Biosocial Theory for Disorders of Overcontrol. The following paragraphs outline the postulates of the two models.

2.5.2 The Neuroregulatory Model of Socioemotional Functioning.

The Neuroregulatory Model of Socioemotional Functioning (Lynch, 2018) is an integrative theory of socioemotional functioning that accounts for how neurobiological mechanisms regulate the social behaviours and emotional responses of individuals. It integrates two models of socioemotional functioning: the top-down model of neurovisceral integration (Thayer & Lane, 2000, 2009), which explains how higher-order brain structures regulate the body’s emotional responses, and the bottom-up polyvagal theory (Porges, 1995, 2001, 2003, 2007, 2009), which explains how different physiological states of the nervous system (and particularly those related to the vagus nerve) influence individual’s responses to social stimuli.

Lynch’s (2018) model posits that affective responses in humans evolved to activate appropriate sympathetic and parasympathetic nervous system responses when triggered by emotionally relevant environmental signals (i.e., of safety, novelty, reward, threat, and overwhelming cues). These neuroregulatory responses function to motivate behaviours of the individual that are perceived to be adaptive in given circumstances, based on the individual’s temperamental tendencies and previous life experiences. In healthy, adaptive, and flexible individuals, these mechanisms function to maintain or restore a state of calm readiness (or social safety), triggering social engagement signals and promoting physiological responses that allow for meaningful

connections with others and open engagement with the environment. As such, the model emphasizes the importance of flexibility and adaptability in responding to an ever-changing environment, facilitated by a well-functioning neurovisceral integration system that modulates autonomic nervous system activity. It also highlights the role of the vagus nerve in achieving and maintaining safety, fostering social engagement and prosocial behaviours, and promoting new learning.

Lynch's (2018) neuroregulatory model also posits that, consequently, lack of adaptability can hinder social connectedness and limit exploration of the environment, eventually leading to adverse outcomes. The second, biosocial model, was put forward as an attempt to explain how inflexibility emerges and persists in the context of maladaptive overcontrol.

2.5.3 The Biosocial Theory for Disorders of Overcontrol.

Building on the neuroregulatory model, Lynch (2018) proposed the Biosocial Theory for Disorders of Overcontrol to explain how maladaptive overcontrol develops and is maintained. He distinguished three categories of factors that, in overcontrolled individuals, result in maladaptive response patterns – nature (biotemperament and genetic predispositions), nurture (sociobiographical factors, including family, cultural, and environmental influences, trauma history, and social learning), and coping (the individual's typical behavioural and emotional responses in challenging circumstances). The following paragraphs consider nature, nurture, and coping factors that interact to reinforce maladaptive overcontrol.

Nature.

The first component of the biosocial theory is nature, which pertains to biotemperament and genetic predispositions. Lynch (2018) proposed several temperamental traits that distinguish biologically overcontrolled individuals. Firstly, similarly to Block (1993), Lynch theorised that overcontrolled individuals are characterised by detail-focused rather than global processing of their environment. He proposed that this makes them more attuned to (even minor) novel cues within the environment. Secondly, Lynch posited that overcontrolled individuals tend to have heightened threat sensitivity, making them more likely to assign negative valence to ambiguous novel cues and perceive them as unsafe – thereby frequently triggering

feelings of anxiety and activating the fight or flight response (or, when overwhelmed, the shutdown freeze response). Thirdly, Lynch proposed that overcontrolled individuals are characterised by low anticipatory reward responses and consummatory reward responses, particularly in social contexts. In other words, they are less sensitive to potentially gratifying social stimuli and require a higher level of prosocial safety signals to both feel desire to affiliate with other people and experience pleasure from such interactions. Further, Lynch theorised that due to the heightened threat sensitivity, overcontrolled individuals are habitually avoidant of novel, ambiguous, and unpredictable situations, and are rigid in their approach to structure, order, and routine. Consequently, this diminishes their capacity for reward-based learning, as such learning is theorised to be most effective when positive reinforcement occurs in response to unpredictable situations (see Hollerman & Schultz, 1998). Lastly, Lynch posited that maladaptively overcontrolled individuals are characterised by excessive inhibitory control. They demonstrate a superior capacity to inhibit behavioural responses to emotional stimuli, control (and inhibit) emotional expression, resist temptation, persist in daunting and distressing tasks, and delay gratification – but at the same time, they are unable to relinquish control even when doing so would be more adaptive. These overcontrolled tendencies related to inhibitory control align with Block’s characterization of overcontrolled individuals.

Interestingly, akin to other prominent models concerned with self-control, Lynch’s (2018) theory overlooks the initiatory component of self-control and does not explicitly address specific tendencies related to initiatory control in maladaptively overcontrolled individuals. Recognising this gap, Hamilton (2021) suggested an extension to the theory that accounts for initiatory control. She proposed that maladaptively overcontrolled individuals exhibit superior initiatory control, allowing them to initiate goal-directed actions even when it is unpleasant and causes them significant distress. The current thesis endorses Hamilton’s positing on the initiatory tendencies of overcontrolled individuals.

Nurture.

The second component of Lynch’s (2018) biosocial theory is nurture, concerned with specific sociobiographical influences that, through bidirectional interaction with nature, can make overcontrol maladaptive, as well as work to maintain and intensify

maladaptive overcontrolled coping. These can be historical or current influences and include family, environmental, and cultural factors, as well as traumatic experiences (e.g., childhood neglect, maltreatment, bullying, and one-off traumatic events, such as natural disasters). In the context of maladaptive overcontrol, these environmental factors may function to reward restrictive and rigid responses (i.e., those described by the four core deficits, such as rigid planning and preparation, dutifulness and diligence, delay of gratification, and quietly enduring distress) while punishing more spontaneous responses that are incongruent with high self-control (e.g., outward expression of emotions or lack of preparation). Perpetual reinforcement of overcontrolled tendencies and punishing of alternative, more flexible ways of behaving can result in a restricted repertoire of over-learned coping strategies, which then negatively impact the individual's social, relational, and occupational functioning. For example, a strict, unsupportive, cold, and critical parent who demands no less than perfection is likely to reward precision, diligence, and constrained emotional expression while harshly punishing even small mistakes, spontaneity, and outward emotional expression. If the child is already biologically predisposed towards overcontrolled responding this can eventually lead to personality dysfunction. Similarly, a culture that equates personal worth with professional success, high productivity, and emotional resilience, may lead to rigid rule-following, constant striving for perfection, and pervasive suppression of emotion, and ultimately result in overcontrol becoming maladaptive over time.

Coping.

The interactions between temperamental predispositions and sociobiographical factors that repeatedly reinforce maladaptive overcontrolled behaviours are posited by Lynch (2018) to lead to the development of a maladaptive overcontrolled coping style. Maladaptive overcontrolled coping is primarily characterised by pervasive masking of feelings, compulsive striving for perfection, avoidance of risk and novelty, overtolerance of distress, and aloof and distant style of relating to others – and manifests through behaviours defined by the four core deficits. As this coping style becomes increasingly rigid, the individual becomes less and less able to adapt their behaviour to changing circumstances and relinquish self-control when appropriate. Additionally, maladaptive overcontrolled coping makes it even more difficult for the individual to attain the state of calm readiness – or, in other words, enter their ‘social safety system.’ This, in turn, results in impaired ability and desire to engage with others

socially and openly interact with one's environment – thereby increasing the feelings of social disconnectedness and limiting the opportunities for new learning.

Maladaptive overcontrolled coping is also posited to be linked to deficits in social signalling. Social signals are defined as communicative or informative actions and behaviours of an individual carried out in the presence of other(s) – whether conscious or unconscious, and verbal or non-verbal (e.g., facial expressions, gestures, head movement, eye gaze, and voice pitch and tone; Lynch, 2018; Vinciarelli et al., 2011). Due to the overactive threat system and underactive social safety system, overcontrolled individuals are likely to present with minimal social signals (e.g., emotionless, blank face, monotonous voice, and lack of gestures) or incongruent (and sometimes overt) social signals (e.g., smiling when angry or upset), as means of suppressing and masking their true emotions. Notably, constrained emotional signalling of overcontrolled individuals does not mean that they do not experience emotions strongly, but rather that they are likely to internalise and discount the importance of emotions to hide what they feel from other people and ensure that they are not perceived as vulnerable.

Literature suggests that ambiguous, inhibited, and incongruent emotional expression in social contexts can disrupt communication, elicit stress responses in people with whom the individual interacts, and make the individual appear untrustworthy (Boone & Buck, 2003; Butler et al., 2003; Mauss et al., 2011). Further, drawing from the work of Sonnby–Borgström (2002) and Sonnby–Borgström and colleagues (2003), Lynch's (2018) theory also posits that constrained social signalling reduces the overcontrolled individual's receptivity to the emotional expressions of others, resulting in impaired empathetic responses. Consequently, deficits in social signalling reduce the overcontrolled individual's chances for developing meaningful social connections and increase the likelihood of aversive and unrewarding social interactions – resulting in social isolation, disconnectedness, and feelings of being different from others (Gilbert et al., 2020; Lynch, 2018). As such, the theory portrays maladaptive overcontrol as primarily a problem of emotional loneliness. Additionally, repeated negative social experiences associated with social signalling deficits further exacerbate the already overt reactivity of the threat system and diminished reactivity of the social safety system, perpetuating maladaptive overcontrolled coping. Lynch

termed this destructive cycle *the kindling effect* to illustrate how repeated exposure to negative stimuli can, counterintuitively, exacerbate the initial problem (see e.g., Barnes & Pines, 2001; Post, 2007).

2.5.4 Integrating Lynch's neuroregulatory and biosocial models.

To summarise the previous paragraphs and integrate the two theories, the Neuroregulatory Model of Socioemotional Functioning accounts for how neurobiological mechanisms influence the socioemotional functioning of individuals, while the Biosocial Theory for Disorders of Overcontrol explains how the bidirectional interplay between three components – nature, nurture, and coping – contributes to the development and maintenance of maladaptive overcontrol. Together, the two models explain how specific overcontrolled biotemperamental predispositions may interact with certain family, environmental, and cultural experiences to foster inflexible, maladaptive coping patterns that limit the individual's capacity for social learning and social connectedness. Over time, through a process of operant conditioning, these maladaptive patterns can become increasingly rigid, making overcontrol maladaptive and eventually resulting in profound emotional loneliness and a range of adverse life outcomes, including complex psychopathology.

2.5.4 Maladaptive overcontrol and links to psychopathology.

As previously mentioned, Lynch (2018) associated maladaptive overcontrol with personality dysfunction. He noted that all personality disorders recognised within the DSM-5 involve difficulties related to self-control, emotion regulation, and interpersonal relationships. He proposed that these difficulties can be further categorised into two overarching classes – undercontrol and overcontrol. Lynch posits that the key characteristics of these classes intersect with two established psychopathological dimensions – internalising and externalising (see Achenbach, 1966; Achenbach et al., 2016). For instance, both maladaptive overcontrol and internalising symptomatology are associated with excessive inhibitory control, emotional constraint, and an aloof interpersonal style. Undercontrol and externalising psychopathology, in turn, are characterised by insufficient inhibitory control, emotional dysregulation, impulsivity, and recklessness. Lynch suggested that maladaptive overcontrol leads to chronic internalising conditions (e.g., cluster A and C personality disorders, anorexia nervosa, autism spectrum disorders [ASD],

obsessive-compulsive symptomatology, and certain psychotic disorders), while maladaptive undercontrol leads to persistent externalising problems (e.g., cluster B personality disorders, attention-deficit/hyperactivity disorder [ADHD], conduct disorders, binge-purge eating disorders, and bipolar disorder).

Interestingly, however, existing literature suggests that externalising and internalising symptoms may coexist (e.g., Essau & de la Torre-Luque, 2023; Levy, 2010; McConaughy & Skiba, 1993; Willner et al., 2016), with some researchers theorising that both internalising and externalising problems may be a result of dysfunction of the same superordinate dimensions of personality or temperament (e.g., Lilienfeld, 2003). Similarly, many clients diagnosed with several mental health disorders exhibit a mix of internalising and externalising symptoms, complicating the diagnostic process, and presenting treatment challenges (e.g., ADHD; Connor et al., 2003; Gnanavel et al., 2019; borderline personality disorder; Choate et al., 2021; Eaton et al., 2011; Ha et al., 2014). In contrast, Lynch's (2018) model is transdiagnostic, and focuses on targeting personality dysfunction. It justly underscores the importance of individual differences between individuals in relation to self-control, emotional regulation, and interpersonal styles of relating to others when issuing diagnoses and planning treatment. It emphasizes that 'one size does not fit all' when it comes to psychological interventions and highlights the significance of considering the personality and unique needs of each client. In this context, Lynch recognised a significant gap in relation to psychological treatments – the lack of interventions appropriate for people characterised by rigid controls and emotional constraint. Underpinned by his neurobiosocial model, he developed RO DBT to target disorders associated with maladaptive overcontrol.

2.5.6 Radically Open Dialectical Behaviour Therapy (RO DBT).

RO DBT (Lynch, 2018) is the first transdiagnostic treatment aimed specifically at treating maladaptive overcontrol, and it is based on the Neurobiosocial Theory for Disorders of Overcontrol. The aim of the treatment is to help overcontrolled individuals relax their self-control and increase emotional expression when adaptive (Luoma et al., 2018). It was designed to target maladaptive neuroceptive tendencies and deficits in social signalling to increase social connectedness, receptivity, openness,

emotional expressivity, and flexible responding of overcontrolled individuals (Lynch, 2018).

RO DBT involves approximately 30 weeks of structured individual and group therapy sessions, complemented by optional telephone sessions scheduled on an as-needed basis. It has been trialled in a variety of settings, including in-patient (e.g., Lynch et al., 2013), outpatient (e.g., Baudinet et al., 2020; Isaksson, Ghaderi, Ramklint, et al., 2021; Lynch et al., 2013), community (e.g., Cornwall et al., 2020), and forensic settings (Hamilton, 2021; Hamilton et al., 2021; Hempel et al., 2018), and has been used with adults (e.g., Cornwall et al., 2020; Isaksson, Ghaderi, Ramklint, et al., 2021; Lynch et al., 2013, 2015) and adolescents (e.g., Baudinet et al., 2021, 2022; Fennig, 2023).

Preliminary evidence from RO DBT evaluation studies indicates that the therapy may be a viable way forward in addressing maladaptive overcontrolled coping in clients with diagnoses of treatment-resistant depression (Lynch et al., 2018), restrictive eating disorders (e.g., Lynch et al., 2013; Isaksson, Ghaderi, Ramklint, et al., 2021; Isaksson, Ghaderi, Wolf-Arehult, Öster, & Ramklint, 2021), perfectionism (Little & Codd III, 2020), and autism spectrum conditions (Cornwall et al., 2020). RO DBT is also being implemented in other clinical and forensic populations (and in various age groups), including with clients diagnosed with post-traumatic stress disorder, several personality disorders, anxiety, alexithymia, as well as with autistic and intellectually disabled clients, and with individuals who engage in non-suicidal self-injury (Hamilton et al., 2021; Hempel et al., 2018; E. Smith-Lynch, personal communication, June 19, 2024; R. Hempel, personal communication, June 19, 2024; R. Hicks, personal communication, June 18, 2024). However, evaluation studies that would assess the effectiveness of RO DBT in these specific populations are yet to be published. Unfortunately, despite the evaluation studies yielding good results, the empirical evidence that supports the Neurobiosocial Theory for Disorders of Overcontrol (Lynch, 2018) remains limited, hindering the implementation of the therapy on an even wider scale.

2.5.7 The evidence gap.

With RO DBT being directly based on the Neurobiosocial Theory for Disorders of Overcontrol, the promising early evidence of the therapy's effectiveness serves as indirect evidence supporting the theoretical framework guiding it. However, direct testing of the model's assumptions – i.e., the posited mechanisms behind the development and maintenance of maladaptive overcontrol and behavioural manifestations of the construct – needs thorough empirical testing if the idea of excessive self-control negatively affecting well-being is to become widely recognised. However, to date, outside of RO DBT evaluation studies, there is a limited number of studies that directly test the assumptions of Lynch's neurobiosocial model.

Evidence for Lynch's theorising in adult populations.

To date, direct testing of Lynch's (2018) theory in adults is limited to a very small number of forensic studies conducted by Hamilton (2021). In her doctoral thesis, she revisited the idea of overcontrol and offending and conducted a systematic review of previous forensic studies in the context of Lynch's (2018) positing on nature, nurture, and coping. In the subsequent original studies, Hamilton focused predominantly on assessing elements of the nurture and coping components of Lynch's biosocial model. She compared clinical, forensic, and socio-developmental characteristics of individuals with convictions diagnosed with personality disorders who were assessed as under- and overcontrolled.

Hamilton's work provided partial support for Lynch's (2018) theorising. Overcontrolled individuals were shown to be highly inhibited, both in terms of resisting impulses, hiding their emotions, and controlling the expression of anger. Their upbringing was typically unstable, and childhood maltreatment, including experiences of emotional and sexual (but less so physical) abuse were common. Overcontrolled individuals were, however, less likely than undercontrolled individuals to act out and have antisocial peers in youth, and more likely to quietly endure the abuse in their home growing up. The findings also supported the tendencies for overcontrolled individuals to be hypervigilant, suspicious, avoidant, dutiful, and rigid in their ways. They were also characterised by high moral certitude, high need for structure and order, and interpersonal aloofness. They reported feeling like an outsider, and different from other people. Additionally, observer-rated scales revealed the tendency for

overcontrolled individuals to lack prosocial signalling, to be more withdrawn, and less spontaneous and fun-loving.

There are, however, some limitations to Hamilton's work. Firstly, the sample was limited to a specialised in-patient population of men with convictions and personality disorder diagnoses, limiting the generalisability of the results. Secondly, the assignment of participants to under- and overcontrolled groups in the original studies was based on clinical sample assignment and pre-existent data from proxy measures, possibly introducing classification bias. Nevertheless, the results of the described studies provided partial proof of concept and preliminary evidence supporting the theory. They also highlighted the need to shift away from perceiving offending as an exclusively undercontrolled problem and highlighted the need for further testing of Lynch's Neurobiosocial Theory for Disorders of Overcontrol in forensic populations and beyond.

Evidence for Lynch's theorising in young people.

Empirical evidence for the Neurobiosocial Theory for Disorders of Overcontrol in non-adult populations is also limited. A study by Gilbert and colleagues (2019) provided support for the presence of several characteristics that Lynch (2018) linked to overcontrolled personality in children aged 4-7. The study's primary goal was validating a newly developed measure for assessing overcontrol in youth – the Overcontrol in Youth Checklist (OCYC). The two-dimensional OCYC was found to have satisfactory psychometric properties. As part of the validation process, several theory assumptions have also been tested, providing preliminary evidence for the relevance of Lynch's framework to the development and functioning of young people.

The results of the study indicated that higher overcontrolled tendencies in children were positively correlated with increased social withdrawal and isolation, poorer social relationships, more experiences of bullying, increased behavioural inhibition and emotional suppression, and high perfectionism. Overcontrolled children were also found to have better reaction times when completing a behavioural task, indicating high engagement with the task. While they initially made more mistakes on the task, they presented with better accuracy on the following attempts, suggesting improved focus and the tendency to strive for better outcomes. Further, the study found

overcontrolled children to present with diminished activation of a neural marker Δ ERN, the underactivation of which was previously linked to higher threat sensitivity, maternal anxiety, and anxiety symptoms. The marker is associated with monitoring of errors, behaviours, and social performance – all characteristic of overcontrol. Overcontrolled children were also found more likely to present with depressive and anxiety symptoms. Later studies using OCYC to screen for overcontrol in children supported the original study's results and found higher overcontrol to be associated with higher inhibitory control, paediatric anxiety, increased obsessive-compulsive symptoms, increased risk for aberrant neural error responding via ERN underactivation, and decreased ability for cognitive shifting (Gilbert et al., 2022; Gilbert, Perino, et al., 2020).

One limitation of the studies is their exclusive focus on children aged 4-7. The assumptions of Lynch's (2018) neurobiosocial model are yet to have been tested in young people of other age groups, except for studies that evaluate RO DBT effectiveness.

The need for more evidence.

To the author's best knowledge, no additional empirical studies have, as of yet, been published that directly attempt to validate the assumptions of Lynch's (2018) theoretical model. For the research on overcontrol to proliferate, and for RO DBT to be offered on a wider scale, there is a dire need to conduct more research on the presentations of maladaptive overcontrol in various contexts and populations, and to verify the claims about the mechanisms behind its development and maintenance. The importance of thorough theory testing in psychological science has been long recognised (Makel et al., 2012; Viswesvaran & Ones, 1995) and appears especially important in the context of clinical psychology and psychiatry. Bauer (2007) emphasised the importance of evidence-based practice in psychology as a healthcare discipline and argued that decisions that can have a significant effect on clients' life and well-being – including those about psychological treatment – should always be grounded in solid empirical evidence. Unfortunately, the advancement of research and, thereby, clinical practice in the context of maladaptive overcontrol in adults is being hindered by a critical challenge – the absence of a quick, valid, and reliable assessment measure that allows researchers to confidently identify maladaptive overcontrol.

2.6 The current protocol for assessing maladaptive overcontrol

Lynch's (2018) theorising regarding overcontrol it is yet to be empirically tested, and there is a dire need for more research so that the idea of maladaptive overcontrol is not brushed to the side yet again. However, as of now, there is no valid, reliable, quick, and cost-effective instrument that would allow researchers and clinicians to confidently identify individuals who are maladaptively overcontrolled. The previous sections of this chapter have discussed how self-control is currently assessed and illustrated that self-control measures aimed at adult populations that demonstrate good psychometric properties fail to appropriately measure maladaptive overcontrol. This section, in turn, critically considers the overcontrol diagnostic protocol proposed by Lynch as part of his theorising.

The current protocol for diagnosing maladaptive overcontrol (Lynch, 2018) comprises the following three steps:

- three self-report questionnaires completed by the assessed individual, which can be a part of a battery of measures that are routinely completed;
- an optional diagnostic clinical interview with the client; and
- three clinician-rated questionnaire(s).

Each of these steps is considered in more detail below, after which the strengths and limitations of the diagnostic protocol are summarised.

2.6.1 Step 1 – the self-report questionnaires.

Assessing Styles of Coping: Word-Pair Checklist (ASC-WP).

When maladaptive overcontrolled coping is suspected, the client is first asked to complete the ASC-WP – a non-clinical measure designed by Lynch (2018) to assess whether individuals lean more towards under- or overcontrolled coping style. It was derived from research by Ashton and colleagues (2004) and Goldberg and Kilkowski (1985) that focused on determining a list of words that are descriptive of personality. The ASC-WP is comprised 47 pairs of words or phrases (e.g., *impulsive – deliberate*), with words in the column A representative of the undercontrolled tendencies (e.g., *impulsive*) and words in column B representative of the overcontrolled tendencies (e.g., *deliberate*). For each pair, the assessed is asked to select the word that is a better description of themselves. A majority of selections in column A indicates a tendency

towards undercontrolled coping, while a greater number of choices in column B suggests a tendency towards overcontrolled coping. Importantly, even a very high number of selections in a single column does not necessarily indicate psychopathology. Nevertheless, the ASC-WP can give the clinician an idea on whether further assessment for maladaptive overcontrol is appropriate, and as such, the measure is used as the first step in the overcontrol diagnostic protocol.

There are, however, several issues with the ASC-WP. Firstly, while a forced-choice approach had been chosen to mitigate social desirability bias (Lynch, 2018), many words in column A (undercontrolled tendencies) appear to have negative connotations (e.g., *naive*, *intoxicated*, *thoughtless*) when compared to the corresponding words within the column B (*worldly*, *clearheaded*, *thoughtful*) – likely making it more socially desirable to select words indicating overcontrolled tendencies. Secondly, the measure appears to be semantically difficult, with many words rarely used in everyday language (e.g., *fastidious*, *impervious*, *shrewd*, *vacillating*, *affable*, *brash*). Semantic difficulty is a recognised problem in psychometric measure development, with items that are difficult to understand suggested to introduce irrelevant variance and increase measurement error (Lenzner, 2012). Lastly, nothing is known about the psychometric properties of the ASC-WP, as, to date, the measure has not undergone psychometric testing. As such, it is unknown whether the scale measures what it claims to measure, and, if so, how accurately. This significantly reduces its utility and puts the reliability of the assessment in question. This said, considering the current lack of alternative options, it can be useful as a preliminary screening measure to assess for overcontrolled tendencies.

Personal Need for Structure (PNS) and Acceptance and Action Questionnaire-II (AAQ-II).

If the ASC-WP indicates high overcontrolled tendencies, two other self-report measures are administered to the client – the PNS, designed by Neuberg and Newsom (1993) to measure individual's preferences for simple structure, and the AAQ-II, designed by Bond and colleagues (2011) to measure psychological inflexibility and experiential avoidance.

The PNS is scored on a scale from 1 (*strongly disagree*) to 6 (*strongly agree*). It consists of two factors: (1) a 4-item *Desire for Structure* (e.g., *I enjoy having a clear and structured mode of life.*), and (2) a 6-item *Response to Lack of Structure* (e.g., *I don't like situations that are uncertain.*). Higher scores indicate a higher desire for structure in Factor 1 and a more negative response to the lack of structure in Factor 2. Newberg and Newsom (1993) found the scale to have satisfactory psychometric properties, with Cronbach's alpha value of .77 and a 12-week test-retest reliability of $r = .84$ for Factor 1 score, $r = .79$ for Factor 2 score, and $r = .76$ for the global score. High scores on the PNS have been linked to certain characteristics that Lynch (2018) associated with maladaptive overcontrol, such as low openness, fixed beliefs, strong focus on performance, dutifulness, a preference for structure and order (including in social settings), and a tendency to apply previously learnt social categories in novel and ambiguous contexts.

For the unidimensional AAQ-II, the items (e.g., *I worry about not being able to control my worries and feelings*) are scored on a scale from 1 (*never true*) to 7 (*always true*), with higher scores being indicative of greater psychological inflexibility. Bond and colleagues (2011) reported the scale as having satisfactory psychometric properties, with Cronbach's alpha values of .78-.88 across six tested samples and a 3-month test-retest reliability of .81 and a 12-month test-retest reliability of .79. Higher levels of psychological inflexibility as indicated by the AAQ-II have been linked to higher levels of depression, anxiety, stress, and shown to predict psychological distress. AAQ-II was previously used to measure psychological inflexibility in RO DBT randomised controlled trials, and increased flexibility post-treatment was found (Lynch et al., 2020). Notably, some researchers have raised concerns over the validity of the AAQ-II and suggested that the scale measures psychological distress rather than inflexibility (see e.g., Tyndall et al., 2019; Wolgast, 2014).

Despite apparent links of the two scales to Lynch's (2018) conceptualisation of overcontrol, it is necessary to keep in mind that neither of the two measures were designed specifically to assess maladaptive overcontrol. As such, each measure only assesses certain elements of the theoretical model. Even when used together, the scales do not measure the full complexity of the construct of maladaptive overcontrol. While they can give the therapist an indication of possible problems associated with

overcontrol, they cannot be used as a standalone assessment. Caution is necessary when interpreting the results.

2.6.2 Step 2 – the diagnostic interview.

Following completion of the three self-report measures by the assessed individual, the clinician may opt to conduct a diagnostic interview with the client. The aim the interview is to assess disorders linked to overcontrol and further identify potential presence of broad features of overcontrol as per Lynch's (2018) theorising. The assessor may choose for the interview to be either structured or unstructured. Lynch (2018) provided example questions that may help facilitate an unstructured interview when assessing for maladaptive overcontrol (e.g., *Are you cautious and careful about how you do things?*; p. 80). As examples of a structured interview that may be particularly helpful, he provided the Structured Clinical Interview for DSM-5 (First et al., 2015), Structured Clinical Interview for DSM-IV Axis II Personality Disorders (First et al., 1997), and the International Personality Disorder Examination (Loranger et al., 1997) – all of which have been shown to have solid psychometric properties.

Although the interview is optional, it is standard practice in most clinical and forensic settings (Lynch, 2018), and utilising interviews in the clinical assessment process is widely endorsed by professionals (see e.g., Maruish, 2014; Sharp et al., 2013), with the multimethod approach argued to provide the most complete picture of client difficulties (Meyer et al., 2001). Importantly, in the context of assessing for maladaptive overcontrol in clinical settings, conducting an interview adds an additional layer of validity, therefore addressing some of the limitations of the Step 1 self-report measures. However, in research settings, having to rely on clinicians to conduct assessments is both costly and time-consuming. As such, while ideal for thorough assessments of psychopathology, such multimethod assessments are not necessarily practical in research contexts – and therefore a comprehensive and reliable psychometric test designed to assess for specific overcontrolled difficulties is crucial when building the empirical evidence base.

2.6.3 Step 3 – clinician-rated questionnaires.

The last part of the overcontrol diagnostic protocol (Lynch, 2018) involves two clinician-rated questionnaires – (1) the Clinician Rated OC Trait Rating Scale and (2)

the Overcontrolled Global Prototype Rating Scale. Both the scales have been developed by Lynch (2018) in accordance with this theorising.

The Clinician Rated OC Trait Rating Scale is a 7-point scale to assess the extent to which the client presents with eight personality traits characteristic of overcontrol according to Lynch's (2018) theorising – openness to experience (reverse-scored), affiliation needs (reverse-scored), trait negative emotionality, trait positive emotionality (reverse-scored), inhibited emotional expressivity, moral certitude, compulsive striving, and high detail-focused processing. The traits are clearly defined and described as a reference for the assessor. According to Lynch, scores 6-7 indicate the presence of maladaptive levels of a trait, while a global score of 40 or above indicates overall presence of maladaptive overcontrolled behaviour. Clinicians are encouraged to use single-trait ratings when determining the urgency of particular treatment targets in RO DBT.

The Overcontrolled Global Prototype Rating Scale is a 5-point Likert-type scale that serves to assess the presence of the four core deficits of maladaptive overcontrol (Lynch, 2018) and is informed by the prototype models of personality assessments proposed by Westen and colleagues (2010). Each of the four core deficits is split into two sections that denote groups of behaviours critical to that deficit, which are defined and described for reference. Scores 4-5 indicate the presence of problematic behaviours belonging to the particular group, while global scores of 17 or higher indicate the presence of maladaptive overcontrol. Of note, for those clients who meet the threshold for maladaptive overcontrol, the clinician completes an additional form scored in the same way – an Overcontrolled Subtype Rating Scale. It describes two subtypes of overcontrolled individuals distinguished by Lynch (2018) – the overly disagreeable and the overly agreeable subtype. The behaviours characteristic to the subtype are described, and their presence is then rated by the clinician. The information on the clients' agreeableness also guides tailored treatment planning in RO DBT, as clients may have different needs depending on their style of relating to others³.

³ The aim of this thesis is to develop an assessment that will identify maladaptive overcontrol in individuals regardless of their agreeableness. Adding an agreeability dimension would add complexity to the conceptual model of maladaptive overcontrol. Given the early stage of research in this area and the resource limitations of the project, incorporating such a complex model is not feasible at this time.

While Lynch (2018) indicated that psychometric properties of the scales have been under evaluation at the time, to the researcher's best knowledge, the results of the evaluation have not yet been published. As such, despite cut-offs being guided by previous research on various disorders, there is no concrete evidence for the validity and reliability of the scales utilised within the diagnostic protocol. A further limitation is that, as with the diagnostic interview, input of a specially trained clinician is necessary, limiting the utility of the scales in research contexts.

2.6.4 The need for a better psychometric measure.

As previously noted, the self-report measures currently utilised within the overcontrol diagnostic protocol (Lynch, 2018) cannot be used as a standalone method to confidently assess individuals for maladaptive overcontrol. However, the proposed follow-up assessments that can increase the validity of the assessment and provide a comprehensive picture of client difficulties require specific clinical training. Many researchers who work in the fields of personality pathology, clinical psychology, and forensic psychology are not fully clinically trained. Further, looking specifically at assessing maladaptive overcontrol, a vast majority of trained clinicians are not trained in overcontrol and RO DBT, due to both the novelty of the theoretical framework and the popularity of the linear view of self-control. As such, the current diagnostic protocol cannot be utilised in research settings to identify overcontrolled participants in a resource-efficient manner. The need for trained clinicians to assess participants for all studies, even non-clinical, increases both the time and costs of the research. In such cases, a robust psychometric measure offers a more feasible alternative for selecting samples for research purposes, as well as clinical screening.

2.7 Rationale

The current chapter highlighted the pressing need for a new, robust psychometric measure that would allow researchers to confidently assess maladaptive overcontrol. Existing measures widely used to assess self-control, such as the BSCS and the LSCS, endorse a linear view of the construct and lack items that accurately capture the theorised characteristics of maladaptive overcontrol. The only measure that endorses the quadratic view of self-control and includes items reflective of overcontrol, the

As such, overly agreeable and overly disagreeable subtypes of overcontrolled individuals are not considered further in this thesis, unless it is crucial to consider them to explain specific findings.

EUC, exhibits poor psychometric properties. A revised version of the EUC, the EUC-13, has also shown some reliability issues and has not been tested in English-speaking populations. Further, the framework underpinning the EUC does not fully align with Lynch's (2018) theorising on maladaptive overcontrol and its deficits. The current diagnostic protocol endorsed by Lynch (2018), while developed in line with his theorising, is lengthy, time-consuming, requires an input from a specially trained clinician, and includes assessments that have not been validated or may not be entirely fit for purpose.

The overarching aim of this thesis is to develop a valid and reliable self-report measure of maladaptive overcontrol that aligns with Lynch's (2018) Neurobiosocial Theory for Disorders of Overcontrol, can be administered efficiently, and does not require the input of a specially trained clinician. Additionally, the proposed assessment aims to evaluate not only the overall extent of maladaptive overcontrol, but also highlight specific overcontrolled deficits that may be most problematic for each individual. To achieve the above aims, a series of rigorous studies was conducted within the general English-speaking population, adhering to best practices for self-report scale development as closely as possible given available resources. The following chapters detail the step-by-step scale development and validation process, discuss relevant conceptual and methodological issues, and finally present the newly developed Overcontrol Assessment Questionnaire (OAQ).

Of note, the primary aim of this project is to produce a scale for research purposes. Nevertheless, it is hoped that, in the future, the scale can also be evaluated and utilised as an assessment tool in various clinical and forensic settings, thereby aiding the clinical assessment process, and increasing access to appropriate treatments tailored to the needs of the clients.

3. Conceptual framework, functional properties, and item generation

3.1 Introduction

Despite a widespread belief that high self-control is adaptive and desirable, a growing body of evidence suggests that excessive self-control – maladaptive overcontrol – can lead to a variety of adverse life outcomes, complex socio-emotional problems, and psychopathology. However, a standardised measure that would allow researchers and clinicians to assess individuals for maladaptive overcontrol quickly and accurately is lacking, hindering progress in the area. The overarching aim of this programme of work is to develop a valid and reliable self-report measure of maladaptive overcontrol.

The primary goal of the scale development process is to produce an instrument that provides meaningful information about the assessed individual's traits and behaviours related to the construct of interest (Colton & Covert, 2007). In the context of psychopathology, psychometric testing is commonly used to guide empirical research that helps to test theoretical models and better understand the specific clinical needs of clients, diagnose disorders, predict the behaviour of individuals in various settings, and ensure that appropriate interventions are offered to clients (Colton & Covert, 2007; Nunnally, 1978; Murphy & Davidshofer, 2005). As such, it is hoped that the new measure of maladaptive overcontrol will encourage further research in the area, so that different elements of the Neurobiosocial Theory for Disorders of Overcontrol can be empirically tested, and the unique needs of overcontrolled individuals' can be better understood. Further, it is hoped that, upon later clinical validation, the scale will also allow clinicians to accurately identify overcontrolled individuals and gain insight regarding the specific nature of their issues, effectively leading to tailored interventions and targeted support being offered on a wider scale.

The current chapter focuses on the first steps in the measure development process. It (1) presents a conceptual framework for the construct of interest, (2) establishes the functional properties of the scale, and (3) presents an initial set of items generated for the new measure.

3.1.1 Conceptual framework.

Psychometric literature indicates that the instrument development process should begin with establishing a clear conceptual framework – i.e., defining the latent construct and its domains and establishing its boundaries (Clark & Watson, 2016). The conceptual framework constitutes the foundation for the later stages of the psychometric process, meaning that an inadequately developed framework is likely to compromise the instrument's psychometric properties (Almanasreh et al., 2019).

In this thesis, the latent construct to be measured by the new instrument is maladaptive overcontrol. As the Neurobiosocial Theory for Disorders of Overcontrol (Lynch, 2018) was the cornerstone of this thesis' understanding of maladaptive overcontrol, the conceptual framework developed directly following the model. Based directly on the model, the researcher defined maladaptive overcontrol as personality dysfunction that stems from bidirectional interactions between specific temperamental predispositions, environmental stimuli and coping mechanisms, and manifests via common maladaptive patterns of behaviours, traits, and maladaptive coping strategies. It is considered to be a manifestation of excessive, inflexible self-control, and distinct from optimal self-control (commonly referred to as 'high' self-control in literature endorsing the linear view). Lynch's (2018) theorising regarding behavioural manifestations of the maladaptive overcontrolled patterns was used to establish the domains and boundaries of the latent construct, the four core deficits of maladaptive overcontrol are operationalised as separate domains of maladaptive overcontrol:

- (1) low flexible control (LFC);
- (2) low receptivity and openness (LRO);
- (3) low social connectedness and intimacy with others (LSC); and
- (4) pervasive inhibited emotional expression and low emotional awareness (PIE).

In this understanding, maladaptive overcontrol is a higher-order factor explained by four lower-order factors (the deficits). A summary of the conceptual framework is presented in Table 3.1. Importantly, maladaptive overcontrol should be differentiated from adaptive overcontrol – a tendency towards overcontrol with the ability to flexibly exert and relinquish self-control depending on the environmental demands.

3.2.2 Functional properties of the scale

Furr (2011) indicated that before items are generated, it is necessary to make decisions regarding the functional properties of the new measure: the scaling method, number of response options, labelling of response options, and mid-point responses. In

Table 3.1: The conceptual framework of maladaptive overcontrol.

Maladaptive overcontrol		
personality dysfunction that stems from bidirectional interactions between specific temperamental predispositions, environmental stimuli, and coping mechanisms, and is manifested through specific common maladaptive patterns of traits, behaviours, and coping strategies		
Deficit	Posited manifestation (Lynch, 2018, p. 8)	Conceptualisation
Low flexible control (LFC)	<i>“manifested by compulsive needs for structure and order; hyperperfectionism; high social obligation and dutifulness; compulsive rehearsal, premeditation, and planning; compulsive fixing and approach coping; rigid rule-governed behavior; high moral certitude (the conviction that there is only one right way of doing something)”</i>	Traits, behaviours, and coping strategies associated with inflexible responding representative of individuals characterised by maladaptive overcontrol
Low receptivity and openness (LRO)	<i>“manifested by low openness to novel, unexpected, or disconfirming feedback; avoidance of uncertainty or unplanned risks; suspiciousness; hypervigilance regarding potential threats; and marked tendencies to discount or dismiss critical feedback”</i>	Traits, behaviours, and coping strategies associated with responses to novel, unexpected, discrepant, and ambiguous stimuli representative of individuals characterised by maladaptive overcontrol
Low social connectedness and intimacy with others (LSC)	<i>“manifested by aloof and distant relationships; a feeling of being different from other people; frequent social comparisons; high envy and bitterness; and reduced empathy”</i>	Traits, behaviours, and coping strategies associated with the typical style of relating to other people representative of individuals characterised by maladaptive overcontrol
Pervasive inhibited emotional expression and low emotional awareness (PIE)	<i>“manifested by context-inappropriate inhibition of emotional expression (for example, presentation of a flat face in response to a compliment) or by insincere or incongruent expressions of emotion (for example, a smile in response to</i>	Traits, behaviours, and coping strategies associated with recognition, processing, and expression (both verbal and non-verbal) of affective states representative of

distress, or a show of concern when no concern is actually felt)” individuals characterised by maladaptive overcontrol

Note. The conceptual framework was derived directly from the Neurobiosocial Theory for Disorders of Overcontrol. Posited manifestations of each deficit are direct quotes from Lynch (2018). The definition of maladaptive overcontrol and conceptualisations of each domain were established based on the model.

this section, different scaling models and approaches to numbering and labelling items are considered, and decisions regarding the intended properties of the new measure made based on the reviewed literature are presented.

3.2.1 Scaling model.

There are three main scaling methods in measure development: Likert-type model (Likert, 1932), Thurstonian model (Thurstone, 1927, 1928, 1929), and semantic-differential model (Osgood et al., 1957). The most widely employed method of scale construction is Likert’s (1932) method, which adopts the *dominance model* of item responding, with the response options normally presented as qualitatively labelled numerical values (Dalal et al., 2014; Stark et al., 2006). For a typical Likert-type scale, higher numerical values represent stronger agreement with positively worded items and thereby, a higher level on the measured trait or behaviour (Dalal et al., 2014). Only items at the extreme ends of the attribute continuum are included in scales that follow the dominance model, with neutral or moderate standpoints being inferred from a mid-point response selection (Dalal et al., 2014).

A common alternative to Likert-type scaling is Thurstone’s (1927, 1928, 1929) method, which follows the *ideal point model* of item responding, in which different items are written to cover the entire continuum of respondents’ attitudes towards the latent construct, including items designed to identify moderate standpoints (Dragow et al., 2010). The model assumes that an individual’s neutral standing on the latent attribute is measured by their response to the moderately-worded items rather than the selection of mid-points (Dalal et al., 2014). Selection of the Thurstonian method appears most appropriate for measures of latent attributes that exist on a continuum and are designed to discriminate between diverse groups of individuals, where both extreme and moderate standpoints are of interest. However, for scales designed to identify extreme or maladaptive levels of a latent attribute (e.g., adaptive fear versus

phobia), Likert's approach appears more suitable. Further, the choice of the ideal point model complicates the item generation and scoring procedures, making the process of scale construction more complex and time-consuming (Barclay & Weaver, 1962). Likert-type scales have also been indicated to produce higher reliability coefficients with fewer scale items when compared to Thurstonian scales, and the methods have been suggested to produce comparable results (Edwards & Kenney, 1946).

Another frequently employed scaling model is the semantic-differential model (Osgood et al., 1957). In this model, for each item on the scale, a pair of adjectives is presented, with each adjective representing an opposite end of the spectrum relevant to the construct of interest (Furr, 2011). Respondents are required to choose the adjective that is more representative of the target of assessment (e.g., in self-report scales – themselves), and the responses are totalled across all pairs to indicate where the target of assessment is situated on the latent attribute spectrum (Furr, 2011). The semantic-differential model, like the Thurstonian model, is therefore a more appropriate choice for scales designed to measure latent attributes that exist on a continuum and are focused on discriminating between dissimilar groups of individuals rather than for scales designed to identify maladaptive levels of a construct.

The aim of this programme of work was to develop a scale that would reliably identify individuals characterised by maladaptive levels of overcontrol, rather than discriminate between overcontrolled, resilient, and undercontrolled individuals. As such, the Likert's scaling model was determined to be the most appropriate.

3.2.2 Mid-point response option.

The dominance scaling model assumes that participants who do hold a moderate standing on the latent construct will select a neutral mid-point response (e.g., *Neither agree nor disagree* or *Neutral*; Furr, 2011; Dalal et al., 2014). However, much discussion has been on whether mid-point options represent a moderate standpoint and should be included in Likert-type scales.

For instance, Nadler and colleagues (2015) indicated that the interpretation of mid-point responses varied substantially between respondents and did not necessarily reflect the respondents' neutral or moderate standing. They argued that the inclusion of a mid-point could increase the likelihood of biased responding. Kulas et al. (2008)

found that correlations of scale scores were nearly identical regardless of whether a mid-point was included, making the response option redundant. Both groups of researchers reasoned against including mid-point response options in psychometric scales. However, both also argued in favour of providing an alternative neutral response option without an assigned numerical value – such as *Not applicable* or *No opinion*. They recommended treating this response option's endorsement as if the respondent skipped the question. While this approach may also allow for an escape from a meaningful answer to an item, the researchers presented evidence that the inclusion of a *No opinion* option is chosen less commonly compared to a mid-point, and that including it may increase the reliability of the scale and decrease the risk of biases (Kulas et al., 2008; Nadler et al., 2015).

In the light of the above argument, a decision was made not to include a mid-point response option with an assigned numerical value in the new scale. Instead, a non-numerical ‘No opinion’ response option was included – the endorsement of which was treated as if the respondent skipped the question.

3.2.3 Number and labelling of response options.

Another issue to consider is the total number of response options. A minimum of two response options is required for a Likert-type scale without a mid-point response option. Nevertheless, a larger number of responses has its advantages. A scale with a larger number of responses may be more sensitive to subtle psychological differences between individuals’ standpoints on the latent construct (Furr, 2011). Lissitz and Green (1975) and Furr (2011) indicated that when too few response options are included, respondents may be forced to choose a response that does not truly reflect their attitude towards the latent construct. However, they also noted that an excessive number of response options may result in increased random error due to indefinable differences between response options.

It has been widely suggested that including four or six response options with assigned numerical values for a scale without mid-points is appropriate (Furr, 2011; Nadler et al., 2015; Simms et al., 2019). Chang (1994) found that 4-point scales may be more reliable than 6-point scales when considering systematic method variance (variance affected by the data collection methods; Nadler et al., 2015). On the contrary,

Simms et al. (2019) recommended 6-point scales over 4-point scales due to replicable, although insignificant, advantages regarding psychometric quality. Further, it has been stressed that labels should signify *equivalent* differences between response options (Furr, 2011). In a typical 4-point scale, response options are *Strongly agree*, *Agree*, *Disagree*, and *Strongly disagree* – and the difference between an endpoint and the next response option (e.g., *Strongly agree* and *Agree*) appears to be more subtle than the difference between the two middle responses (e.g., *Agree* and *Disagree*). Therefore, including two additional response options (e.g., *Somewhat/Slightly agree* and *Somewhat/Slightly disagree*) may be reasonable. While there have been concerns that assigning labels that represent equivalent differences between response options may become more complex as the number of total response options increases above six or seven (Furr, 2011), this was not an issue for scales with seven or fewer response options (e.g., Simms et al., 2019).

Lastly, whether all response options should be qualitatively labelled has also been debated. Many scales only provide labels for endpoint response options but there is evidence suggesting that labelling all response options results in better psychometric quality of the scale (Krosnick et al., 2005).

For the current project, a decision was made to include six response options with assigned numerical values, with each option clearly labelled (1 – *Strongly disagree*, 2 – *Disagree*, 3 – *Slightly disagree*, 4 – *Slightly agree*, 5 – *Agree*, 6 – *Strongly agree*). The researcher also ensured that the generated items were worded to allow for consistent labelling across all the items.

3.3 Item development

Once a conceptual framework has been developed and decisions have been made regarding the functional properties of the scale, the next step is to generate a large item pool that can be reduced, refined, and validated during the later stages of the process (Furr, 2011). It has been suggested that the initial item pool should contain *at least* twice as many items as the intended number of items in the final version of the instrument (Boateng et al., 2018).

Researchers previously indicated that the best practice for developing an initial item pool for a new instrument is to converge deductive and inductive methods for

item generation (e.g., Boateng et al., 2018; DeVellis & Thorpe, 2021; Furr, 2011; Morgado et al., 2017). Deductive methods include conducting a literature review and/or examining existing scales that may be related to the construct of interest. Inductive methods involve exploratory qualitative studies such as individual interviews or focus group discussions with experts on the subject or a sample of participants from the target population. According to Boateng and colleagues (2018), deductive methods provide the necessary theoretical foundations for item generation, while inductive methods allow for a transition from abstract theorising to identifying how the construct of interest is manifested. As such, both deductive (literature review) and inductive (focus groups with the target population) methods for item generation were utilised in the current study.

3.4 Deductive item generation.

Constituting the foundation for the measure development process in the current thesis, the Neurobiosocial Theory for Disorders of Overcontrol (Lynch, 2018) was thoroughly reviewed. Additionally, available literature pertaining to the theory and the RO DBT intervention (Lynch, 2018) developed based on the model was reviewed. Items for the scale were generated based on the literature, and no participants were involved at this stage.

3.5 Inductive item generation.

The study plan was submitted for review by the Schools of Business, Law and Social Sciences Research Ethics Committee (reference ID: 1536443). The application included documents such as participant information sheet, consent forms, and debrief forms – all of which were developed in accordance with the British Psychological Society guidelines and internal Nottingham Trent University guidelines. A favourable ethical opinion was granted for the conduct of the study.

3.5.1 Pilot study.

Design.

The inductive stage of item generation involved conducting an exploratory, mixed-methods study that considered target population views and experiences related

to overcontrol. A two-part online pilot study was conducted to test the intended methods.

Part 1. A non-clinical self-report measure, the ASC-WP, designed to screen for under- or overcontrolled tendencies was administered to participants online to initially screen for participants exhibiting high levels of overcontrolled or undercontrolled tendencies.

Part 2. Focus groups were utilised in this part of the study. A group of participants were invited to participate an online focus group discussion based on their scores on the screening questionnaire and availability. The aim of the discussion was to identify traits and behaviours that can help identify maladaptive overcontrol and distinguish it from both adaptive overcontrol and undercontrol.

Based on ASC-WP scores, the researcher ensured that more participants in the focus group leaned towards overcontrol than undercontrol, due to the intended focus of the measure being the identification of maladaptive overcontrol. However, the researcher decided that inclusion of undercontrolled participants may be useful in terms of stimulating the discussion and identifying important differences between the two personality styles. It was hoped that this would help the researcher develop items that would better capture traits and behaviours unique to overcontrol, especially if the groups experienced similar issues that may have had different underlying causes. For example, it could be that both groups experienced profound loneliness, but only in the overcontrolled group these were linked to the specific biases and deficits described by Lynch (2018) in the theoretical model.

Materials.

Part 1. The ASC-WP (Lynch, 2018) was administered via Qualtrics to participants who expressed an interest in taking part in the focus group. The purpose of the administration of ASC-WP was to screen for participants high in overcontrolled traits and behaviours or high in undercontrolled traits or behaviours. ASC-WP is the first of the three self-report questionnaires administered as part of the current overcontrol diagnostic protocol (Lynch, 2018). It requires participants to read 47 pairs of words (or phrases), and for each pair, to tick a box next to the one they consider to be more descriptive of them. Words and phrases in column A are more characteristic

of undercontrolled traits and coping, while words in column B are more characteristic of overcontrolled traits and coping. Example item from column A is “immediate gratification,” and the corresponding item from column B is “delay gratification.” According to Lynch (2018), the questionnaire “should not be considered a measure of psychopathology but rather a measure of the extent to which a person leans toward an overcontrolled or undercontrolled personality style and style of coping” (p. 77). It is therefore appropriate for use in non-clinical settings. Due to some of the words being uncommon in everyday language, participants were instructed to use Google or a dictionary if unsure of the meaning of some words. Permission for use of the questionnaire was granted by one of the senior managers of Radically Open Ltd.⁴

A demographic questionnaire was also included, with questions regarding the following variables: age, country of origin, country of residence, ethnicity, sex assigned at birth, gender identity, marital status, and educational level. It also included a question on whether English is a participant’s first or preferred language. Doodle polls were used to collect information about availability of participants qualifying for the focus group on selected times. Qualtrics was used to present the participant information sheet and collect consent from participants invited to take part in the focus groups.

Part 2. Microsoft Teams software was used to conduct the focus group discussion, with video- and audio-recording and auto-generated transcription enabled. A semi-structured discussion schedule (Appendix A) was developed by the researcher based on Lynch’s (2018) theorising about core deficits of overcontrol and wider literature on defining the construct of self-control. The schedule was split into five parts. The first part referred to the participants’ understanding of self-control and its three main components (inhibition, initiation, and effort), including questions such as “*How easy would you say resisting temptations and impulses is for you?*”. The next four parts included questions regarding the four core deficits of maladaptive

⁴ Radically Open Ltd. is a RO DBT training and supervision company led by, among others, Thomas Lynch – the developer of the Neurobiosocial Theory for Disorders of Overcontrol, RO DBT, and the current overcontrol diagnostic protocol, including the ASC-WP.

overcontrol (Lynch, 2018), with one section per deficit. Each part included prompts to think about behaviours and responses associated with the specific deficit, guided by the literature review previously conducted by the researcher. For instance, for the low receptivity and openness deficits, one of the prompts prepared was: *“Think of a time when you were faced with uncertainty. If comfortable, describe the situation to others. How did you manage this situation? How did the uncertainty make you feel?”*. To prepare for a situation where neither of the participants could think of a situation relevant to the question or none of them would decide to share their experience, example scenarios were prepared (e.g., *“You did not have time to study for an important exam at university but had to take it anyway. You are not sure how you did, and it is important that you pass. You will not find out the results for another month. How does the uncertainty of this situation make you feel?”*).

Participants.

Part 1. A convenience sampling method was used, with participants recruited via an advert on the researcher’s social media profile with privacy set to ‘Friends only’ or directly approached by the researcher via private messaging. Inclusion criteria were to be over 18 years of age and speak fluent English.

Forty participants completed the pilot screening survey. Thirty-one participants with over 50% of responses indicating overcontrolled personality/coping style were assigned to the overcontrolled group (OC group), and nine participants with over 50% of responses indicating undercontrolled personality/coping style were assigned to the undercontrolled group (UC group; see Table 3.2).

Table 3.2: ASC-WP scores of participants who completed the pilot screening survey.

Variable		OC group N = 31 (77.50%)	UC group N = 9 (22.50%)
Responses in the dominant column (%)	M	66.78	68.25
	SD	10.66	10.31
	Min.	51.06	51.06
	Max.	91.49	78.72

Part 2. To identify participants most likely to exhibit high levels of overcontrol and undercontrol, a mean split was performed separately for the OC and the UC group. Participants scoring above the mean within their respective groups ($N_{\text{total}} = 20$, 50.00%;

$N_{OC} = 16, 51.61\%$; $N_{UC} = 4, 44.44\%$) were invited to complete a poll to indicate their availability for the focus group. Given that the ASC-WP questionnaire has not been previously validated and does not provide clinical thresholds, this methodological strategy was applied to reduce the likelihood of focus group participants representing individuals with lower or moderate levels of the respective self-control tendencies. Since the distribution of scores was skewed toward overcontrol, the mean split allowed for a higher threshold for inclusion in the OC group compared to a median split. While UC participants were also of interest to the study, the primary focus was on recruiting participants with pronounced overcontrolled tendencies.

It was intended that five to seven participants from the overcontrolled group and three to four participants from the undercontrolled group would take part in the discussion, with a total of nine to ten participants. The initial responses to the poll indicated that there were not enough participants with undercontrolled tendencies available to take part at the same time. As such, a decision was made to invite every participant whose scores indicated undercontrolled tendencies, and as such, five additional invitations were sent. Twenty-one of the 25 participants who were asked to provide availability responded to the poll. Three out of eight available times with at least three participants from the undercontrolled group and at least six participants from the overcontrolled group were considered. For each of these times, for both the overcontrolled and undercontrol groups, mean percentage of responses in their dominant column was calculated. Finally, the time with the highest mean percentage of responses in the dominant column for both groups was chosen for the focus group discussion ($M_{OC\ group} = 76.89\%$, $SD = 8.15$, $range = 68.09-91.49\%$; $M_{UC\ group} = 70.21\%$, $SD = 11.26$, $range = 57.45-78.72\%$).

All participants who indicated their availability for the time chosen for the focus group ($N = 9$) were invited to participate. All invited participants consented to taking part in the focus group, however, two did not join the meeting on the day. The final pilot focus group sample (Table 3.3) consisted of three female and four male participants ($M_{age} = 29.86$ years, $SD = 9.63$, $range = 23-51$ years old). All participants were currently residing in the UK, with six participants born in the UK (all of whom indicated English as their first language) and one born in Bulgaria (who indicated

English as their preferred language). For more detailed demographic information about the pilot focus group sample, see Appendix B.

Table 3.3: ASC-WP scores of participants in the pilot focus group.

Variable		OC group N = 5 (57.14%)	UC group N = 2 (28.57%)
Responses in the dominant column (%)	M	76.17	68.09
	SD	9.08	15.04
	Min.	68.08	57.45
	Max.	91.49	78.72

Procedure.

Part 1. Individuals interested in taking part and meeting the above inclusion criteria were asked to fill in an online screening survey. Within the screening survey, participants were first presented with a participant information sheet explaining the purpose of the study, inclusion criteria, withdrawal rights both during and after the study, and an explanation of how the data would be stored and used. To proceed, participants were asked to sign a consent form, where they would confirm that they met the inclusion criteria, that they were happy for the data to be used for research purposes, and that they were happy to be contacted by the researcher if selected for the focus group interview. They were also asked to provide a unique identifier that would allow them to withdraw their data following the completion of the study, as well as to provide an email address so that the researcher could contact them if they were selected to participate in the focus group discussion.

Upon agreeing to participate, participants were presented with the demographic questionnaire. It was explained to participants that they were being asked for the demographic information to help the researcher make sense of the data, and they did not need to disclose any information they did not wish to disclose. Subsequently, the participants were presented with the ASC-WP questionnaire to screen for overcontrolled and undercontrolled personality/coping style. Participants were instructed, for each pair, to select the word or phrase that they considered to be more characteristic of them. They were also advised not to overthink the answers and to go with their intuition. Additionally, participants were advised that if they were unfamiliar with one of the words or phrases on the scale, they could search for the meaning of them online or use a dictionary. After completing the ASC-WP, participants were

shown a debrief screen where they were thanked for participation, reminded about the withdrawal rights, and provided with contact details for the researcher, the researcher's supervisory team, and points of support they could contact if the study had caused them any distress.

Part 2. A sample of participants selected based on their scores on the ASC-WP was invited to complete a poll to indicate their availability for the focus group, with eight available dates. Participants who indicated availability on the time chosen for the focus group based on mean scores were invited to the focus group meeting. The invitation included attached participant information sheet and consent form, questions planned for the discussion, and a link to the online meeting with instructions on how to join. Participants were advised to use their unique ID instead of their name when joining the meeting. Following the start of the meeting, before the recording has commenced, the purpose of the study as well as the procedure were explained. Then, the researcher briefly introduced participants to the concepts of overcontrol and undercontrol. Participants were given a chance to ask questions, following which the recording was started and the discussion began. The discussion schedule was used as scaffolding, while the participants were encouraged to freely discuss the topic and share and compare their experiences. After two hours have passed, participants were asked if they had any questions or would like to add anything or ask any questions. They were thanked for participation, and informed that they would receive a virtual £10 Amazon voucher as a thank you, along with a debrief sheet directing to support points in case they were distressed about anything they had talked about during the study. Participants were also asked to respect each other's confidentiality and avoid discussing anything that emerged during the focus group with others.

3.5.1 Main study.

Design.

Like the pilot study, the main study comprised two consecutive parts: a screening of participants from the target population using the ASC-WP questionnaire and focus group discussions. However, following the completion of the pilot study, the advantages and the challenges of the methods applied were reviewed, and several changes to the design were applied before the main study commenced.

Part 1. Several changes were made to three demographic questions:

- Sexual orientation: the category “Bisexual” was replaced by “Plurisexual (e.g., bisexual, pansexual, fluid)” – an umbrella term for individuals attracted to more than one gender that emerged in sexuality research (Galupo et al., 2017; House et al., 2022).
- Relationship status: a category “In a relationship but not living together” category was added due to several pilot participants who had selected “Other” specifying the category.
- Ethnicity: “Middle Eastern/North African/Arab” category was added as a separate category, as recommended by Maghbouleh et al. (2022). The category name used covers both UK and US Census terminology.

Part 2. Several changes were made to the qualitative part of the study. Firstly, it was decided only to invite participants with overcontrolled tendencies, rather than both under- and overcontrolled. The initial decision to include both undercontrolled and overcontrolled participants was initially made so that differences between how overcontrol and undercontrol manifest might become apparent. However, during the pilot discussion, the researcher was able to observe differences between participants within the overcontrolled group, potentially reflecting and contrasting adaptive and maladaptive overcontrolled behaviours. Exploring the differences between overcontrolled participants was a greater priority than exploring differences between over- and undercontrolled participants, as the researcher aimed for the new measure to be able to discriminate between adaptive and maladaptive levels of overcontrol. It was hoped that including only overcontrolled participants in further focus groups would allow more space to freely discuss and compare their experiences, differences, and similarities in more detail than was possible when undercontrolled participants were also present. Moreover, the analyses revealed that over three quarters of participants who completed the screening leaned towards overcontrolled tendencies, and participants in the pilot disclosed that they were more likely to choose words indicating overcontrolled tendencies due to many words indicating undercontrolled tendencies appearing to have more negative connotations. These results could indicate that the ASC-WP questionnaire used for screening was somewhat biased towards overcontrolled tendencies, with the scale likely being less reliable in identifying

participants with undercontrolled tendencies. Therefore, the researcher felt that focusing on overcontrolled participants in the focus groups that followed was reasonable.

Further, the intended number of participants was reduced to five per focus group. This was based on the experience of the pilot focus group, where the larger group size only allowed for limited content to be discussed during the two hours. The researcher felt that an online discussion between more than five participants would be difficult to manage. It was anticipated that a smaller group size would reduce the time participants needed to commit to the study by allowing for more content to be covered, yet still allow for full engagement of all participants. To additionally ensure that all four domains of overcontrol were sufficiently covered in the discussion, the introductory questions about defining self-control and its components were omitted. Instead, the three main components of self-control were briefly introduced to participants at the beginning of the focus group.

Lastly, a decision was made to inform focus group participants that they were more inclined towards overcontrolled tendencies based on their scores on the screening questionnaire. This was due to several participants in the pilot asking about their scores and being unsure of where they lie on the spectrum. While this appeared quite useful in stimulating the discussion where both undercontrolled and overcontrolled participants were involved, it was anticipated that informing participants would work better in a situation where all of them exhibited relatively pronounced overcontrolled tendencies. For ethical reasons, it was emphasised to participants that the measure used for screening was a non-clinical measure and that the score did not imply maladaptive levels of overcontrol. It was also stressed that the researcher was not qualified to diagnose any conditions, and that in case any worries arose following the discussion, points of support could be found via links provided in the debrief sheet.

Despite these challenges experienced during the pilot study, and the consequent changes in the methods applied during the main study, a decision was made that the data collected during the pilot study would be used alongside the data collected during the main study, as the changes made were not of a nature to preclude incorporating the pilot data.

Materials.

Part 1. Materials used for this part were the same as those used in the pilot study. Please refer to the pilot study materials section for a description.

Part 2. Materials used for the second part were largely the same as those used in the pilot study (refer to the Materials section of the pilot study for description). However, a change was applied to the discussion schedule, with the section of questions related to the definition of self-control and its components (inhibition, initiation, and effort) abandoned. This was for two reasons. Firstly, the interview schedule needed to be shortened, so that all questions could be covered in each focus group within the two-hour time limit. Secondly, during the pilot, the in-depth discussion on components of self-control seemingly directed participants within the pilot to think about overcontrol as high-trait self-control rather than a personality style that could manifest in various behaviours that may not be related to self-control at first glance (e.g., dislike for uncertainty). It was decided that the researcher would simply ask participants to keep in mind that the examples they used during the discussion could include both initiatory and inhibitory actions (with both initiation and inhibition shortly explained), and to, where relevant, consider whether the behaviours they recall require a lot of effort from them.

Participants.

Part 1. The study was advertised via posters placed at Nottingham Trent University, as well online advertisements posted on social media platforms such as Facebook, Twitter, Reddit, and LinkedIn. Individuals were invited to scan a QR code or use a link to complete a screening questionnaire (ASC-WP and a demographic questionnaire) for a chance to qualify to take part in the focus group section of the study. Inclusion criteria were to be over 18 years of age and to speak fluent English.

One hundred and sixty-two responses to the screening survey were submitted. Twenty-five responses were excluded from analyses, with 22 deleted due to a majority of responses to the questionnaire missing, and three responses deleted due to no e-mail address ($N = 2$) or a fictional e-mail address ($N = 1$) provided. One hundred and thirty-five participants provided a response to all questions. For two participants, one answer in the ASC-WP was missing, but they were still included in the analyses. The final

sample consisted of 137 participants, with 23 participants with over 50% of responses in the Column A assigned to the UC group, and 114 participants with over 50% of responses in the Column B assigned to the OC group (see Table 3.4).

Table 3.4: ASC-WP scores of participants who completed the main screening survey.

Variable		OC group N = 114 (82.21%)	UC group N = 23 (16.79%)
Responses in the dominant column (%)	M	72.68	63.61
	SD	11.26	11.22
	Min.	53.19	51.06
	Max.	97.87	87.23

Part 2. Fifty-two participants from the overcontrolled group who scored above their group's mean were sent e-mail invitations to indicate their availability via a poll with 19 available times, ranging from morning to evening to account for potential time zone differences. 41 responses to the poll were recorded. Three responses were deleted as the e-mail address of the poll respondent did not match any of the e-mail addresses of participants who completed the screening survey, leaving 38 valid participant responses ($M_{OC \text{ responses}} = 82.42\%$, $SD = 6.27$, range = 74.47-97.87%). Four times with the highest participant mean scores were chosen for the focus groups, and within that, five people with the highest scores available at each time were invited to participate in the study ($M_{OC \text{ responses}} = 85.53\%$, $SD = 6.35$, range = 76.60-97.87%).

Thirteen of the 20 invited participants returned signed consent forms, with two participants not attending the discussion later. The final post-pilot focus group sample (see Table 3.5 for ASC-WP scores) consisted of four female and seven male participants ($M_{age} = 27.64$ years, $SD = 5.55$, range = 22-43 years old). Nine participants were born and currently residing in the UK (all of whom indicated English as their first language), and two were born and currently resided outside the UK (both of whom indicated English as their preferred language). For more detailed demographic information about post-pilot focus groups participants, see Appendix C.

Table 3.5: ASC-WP scores of post-pilot focus group participants.

Variable		Total (N = 11)	Group 1 (N = 3)	Group 2 (N = 2)	Group 3 (N = 3)	Group 4 (N = 3)
Responses indicating overcontrolled tendencies (%)	M	84.53	78.72	95.74	85.11	82.27
	SD	6.46	2.13	3.01	2.13	3.25
	Min.	76.60	76.60	93.62	82.98	78.72
	Max.	97.87	80.85	97.87	87.23	85.11

Procedure.

Part 1. The screening part of the main study followed the same procedure as in the pilot study. Please refer to the Procedure section of the pilot study for description.

Part 2. Participants who scored above the mean for the overcontrolled group were sent a doodle poll via e-mail and asked to select as many suitable times as possible. They were also informed that they must respond to the poll using the same e-mail address as they used when completing the screening poll so that their identity can be verified.

Poll respondents were assigned to the focus groups based on their scores and availability. For each time, five individuals with the highest scores were selected, and a mean score for the group was calculated. Where a participant was already assigned to one focus group with a higher mean score, their scores were not taken into consideration for any other focus group, and they were replaced with the next highest scoring person. Where there was only one space left for a given focus group, and there was more than one participant with the same score, the participant was selected at random. Selected participants were invited to participate in the focus groups via e-mail, with participant information sheet, consent form, discussion schedule, and instructions on how to join the meeting attached. They were informed they may use a fake name to join if they wish, but their identity may have to be confirmed on the day by the researcher though asking to confirm their initials.

On the agreed date, before the recording has commenced, the purpose of the study as well as the procedure were explained, and boundaries of confidentiality were discussed. The researcher briefly introduced participants to the concepts of overcontrol and undercontrol. Participants were given a chance to ask any questions, following which the recording was started and the discussion began. The discussion schedule was used as a scaffolding, while the participants were encouraged to freely discuss the topic and share and compare their experiences. Questions regarding all four core deficits of maladaptive overcontrol were covered in each of the focus groups, with each discussion lasting between one and a half to two hours. After the focus group,

participants were sent an e-mail with the debrief form and the incentive for participation.

3.6 Data collation and results

In the deductive part, the item pool was generated based on the literature review. In the inductive part, the researcher listened to the recording of each of the five focus group discussions (one from the pilot study and four from the main study), with an automatically generated transcript at hand. Based on experiences shared by the participants during the focus group, the researcher generated a list of potential items to include within the initial item pool. The list was reviewed against the conceptual framework and the items already developed during the deductive stage. Items written based on the focus group content that were within the scope of the conceptual framework and were not repeats of the previously developed items were added to the item pool.

Subsequently, each item for the initial item pool developed via deductive and inductive methods was assigned to one of the four intended subscales (LFC, LRO, LSC, or PIE – set out by the four core deficits of maladaptive overcontrol). The researcher made a conscious effort for the items not to be associated with more than one of the deficits, to decrease the chance for a single item to load onto more than one factor. Where two or more items appeared to have a similar meaning, all were included within the scale, with the assumption that the optimal item would emerge during later stages of the project.

A total of 190 items were generated during the item generation stage (including 31 reverse-scored), with 56 items in the LFC domain, 41 items in the LRO domain, 42 items in the LSC domain, and 51 items in the PIE domain. Tables 3.6-3.9 present the initial item pool for each of the four domains, including information about where each item was derived from – an ‘LR’ label if the item was generated during the deductive stage ($N = 52$ across the four domains), an ‘FG’ label if the item was generated during the inductive stage ($N = 62$ across the four domains), or an ‘LR/FG’ label if the item was written in the deductive stage and endorsed during the inductive stage ($N = 76$ across the four domains).

Table 3.6: The initial item pool for the LFC domain of maladaptive overcontrol.

Item	Source
I like to be prepared for any possibility so that nothing can go wrong.	FG
I like to make detailed plans for everything.	LR/FG
I stick to the detailed plans I make.	LR/FG
You can never be too prepared.	LR/FG
I feel the need to fix all problems immediately.	LR
I would only break a rule if I knew it was the right thing to do.	LR/FG
Breaking rules is not worth the consequences.	FG
I notice small changes that other people may not see.	LR
I have strong opinions about how things should be done.	LR/FG
I strongly dislike it when plans change unexpectedly.	LR/FG
Small changes to my daily routine upset me.	FG
Lack of structure and order upsets me.	LR/FG
I must be the best at everything I do.	LR/FG
I always obsess over small details.	LR/FG
I must be fair, even if I don't want to at times.	LR
When I do someone a favour, I expect that they will also help me when I need it.	FG
I rehearse what I want to say over and over again.	LR/FG
I persist at tasks even if they cause me distress.	LR
Persisting at difficult tasks makes me feel worthwhile.	FG
I like to arrange things in an ordered manner (for example, the items on my desk).	LR
I feel compelled to repeatedly check everything (for example, that the doors are locked).	LR
Being correct is more important to me than it is to most people.	LR
I like my life well-structured and predictable.	LR/FG
I often feel a strong urge to fix things immediately (e.g., arising problems or tensions in relationships).	LR
Asking for help makes me feel like I am not good enough.	LR/FG
I obsessively check the correctness of my work.	LR/FG
Striving for perfection in everything I do makes me feel worthwhile.	LR/FG
*I find it extremely difficult to stick to the plans I have made.	LR/FG
I frequently worry that others see me as incompetent.	LR/FG
I would only ever volunteer to answer a question if I was certain I knew the correct answer.	FG
I frequently wonder if I am working hard enough.	LR/FG
Striving to be the best is more important to me than it is to most people.	FG
I prefer to know what to expect from a situation in advance.	LR/FG
When I look at someone else's work, I frequently think that I could do better than they did.	FG
It's important to me that people adhere to social norms and standards.	LR/FG
Having self-control is more important to me than it is to most people.	LR
Even when a situation requires me to break a rule, I feel guilty about it.	FG
I feel more guilty about breaking rules than most people.	FG
I find it difficult to cope with unexpected changes.	LR/FG
Exact execution of a plan assures good quality of work.	FG
I must always do what I believe is right.	LR
Sometimes I feel like I am out of control even when other people think I am in complete control.	FG
I pride myself in always appearing to be in complete control.	LR/FG
I not only strive to be perfect, but also try to make it look easy.	FG
I never do things at the last minute.	FG
Breaking bad habits requires little effort.	FG
Sticking to a sequence (e.g., morning routine) requires little effort from me.	FG
*Things that seem to come easily to many people require a lot of effort from me.	FG
*I often struggle to stick to a simple routine.	FG
It takes very little effort for me to resist an impulse.	LR/FG
*I easily give in to temptations.	LR/FG
I always stick to the plans I have made.	FG
I cannot stop thinking about a problem until I find a solution.	LR/FG
When someone else tells me they have a problem, I immediately try to find a solution to that problem.	FG
Responsibilities should always come first; having fun can wait.	FG
Doing things to achieve an important goal do not require much effort from me.	FG

Note. Reverse-scored items are indicated with an asterisk (*).

Table 3.7: The initial item pool for the LRO domain of maladaptive overcontrol.

Item	Source
I carefully consider potential consequences before taking a risk.	FG
*I welcome critical feedback even if I did not ask for it.	LR/FG
I purposefully avoid situations in which I could be criticised.	LR
I purposefully avoid situations in which I could be seen as weak or incompetent.	FG
When people criticise me, it is because they want to be like me.	LR/FG
*I often do things just for fun.	LR
I am always on the lookout for potential threats.	LR
*I am always willing to spontaneously try new things.	LR
I am always on high alert.	LR/FG
When somebody gets under my skin, I often contemplate evening the score.	LR
I strongly believe in the saying "better safe than sorry".	LR/FG
There is more to learn from what went wrong than from what went right.	LR
I have been told by other people that I hold myself up to rigorous standards.	LR/FG
I am critical of other people.	LR
I am my own best critic.	LR
*Unpredictability is exciting.	LR/FG
The world is a scary place.	LR
When I make a mistake in front of others, I feel more embarrassed than most people would.	FG
*If I have had a difficult day, I reward myself with something nice.	LR
I rarely feel like I have earned a reward.	LR/FG
Avoiding negative consequences is a better motivator than gaining something positive.	FG
I always weigh up costs and benefits before I do something.	FG
I am only willing to put energy into things that are important to me.	FG
I often see a threat where others may not.	FG
I dislike uncertainty more than most people.	LR/FG
*Life without risks is no fun.	LR/FG
I often downplay my achievements.	FG
I only welcome critical feedback when I have asked for it.	FG
*I am always happy to get out of my comfort zone and try something new.	LR/FG
I need proof that something will happen before I get excited about it.	FG
Even when I get praised for my work, I still think I could have done it better.	FG
I only ever allow myself to be spontaneous when the situation is right (e.g., on holiday).	FG
*I would do anything to avoid being the centre of attention.	FG
I see most things as a potential threat.	FG
I will try to convince people that I'm right even if I know I'm not.	FG
Other people doing better than me makes me feel bad about myself.	FG
*I love trying new dishes in restaurants.	FG
*I love being the centre of attention, whether it is positive or negative.	LR/FG
*I seek out excitement.	LR
Being called out on a mistake in front of other people is my worst fear.	FG
Embarrassing myself in front of other people is my worst fear.	FG

Note. Reverse-scored items are indicated with an asterisk (*).

Table 3.8: The initial item pool for the LSC domain of maladaptive overcontrol.

Item	Source
In social situations, I prefer when people have clearly assigned roles and stick to them.	LR
I rarely feel connected to other people.	LR
I have few close friends, if any.	LR/FG
I feel that I am an outsider.	LR
I feel disconnected from the world.	LR
I feel different from other people.	LR
I feel distant from other people.	LR
*Being around other people makes me feel alive.	LR
*I am a very sociable person.	LR

I generally prefer being on my own rather than around other people.	LR
I dislike pointless social interactions.	LR
I often yearn to be alone for a while after participating in social events.	LR/FG
When disagreements arise, I withdraw from the situation.	LR/FG
It takes me a long time to warm up to other people.	LR/FG
People must prove themselves to me before we can be friends.	FG
I dislike parties.	LR/FG
*I love meeting new people.	LR/FG
*I don't mind starting a conversation with a new person.	FG
I enjoy being difficult to get to know.	LR
I often think others have unfair advantages in life.	LR
I often struggle to understand another person's perspective.	LR
I often feel misunderstood by other people.	LR
*Being around other people helps me relax.	FG
*I am a warm and affectionate person.	LR/FG
People tell me I'm difficult to read.	LR/FG
*My face is very expressive.	LR/FG
*I use gestures a lot when communicating with others.	LR
I have been told I do not smile often.	FG
Strangers often approach me or smile at me in the street.	LR
Comparing myself to people who are worse off often makes me feel better about myself.	LR/FG
I envy people who do better than me.	LR/FG
I compare myself to others without even realising it.	FG
*I am comfortable sitting with my emotions and trying to understand them.	FG
I only enjoy conversations with people that I connect with on an intellectual level.	FG
I strongly dislike small talk.	FG
I only ever cry on my own or in front of people that I am very close to.	FG
My relationships with other people are rather superficial.	LR/FG
I sometimes give others advice even if they don't ask for it.	LR
I can rarely relax in social situations.	LR
I am more awkward in social situations than most people.	LR/FG
I tend to hold grudges for a long time.	LR
I struggle to believe in real, unconditional love.	LR

Note. Reverse-scored items are indicated with an asterisk (*).

Table 3.9: The initial item pool for the PIE domain of maladaptive overcontrol.

Item	Source
I am proud of how well I tolerate distress.	LR/FG
I rarely get very excited.	LR/FG
I get anxious easily.	LR/FG
I only ever feel emotions if I allow myself to.	FG
I can "turn on" and "turn off" my emotions how I please.	FG
I tend to discount and "push down" my emotions.	LR/FG
Being emotional is a sign of weakness.	LR
I never let my struggles show.	LR/FG
Thinking about feeling emotions makes me feel uneasy.	FG
I strongly dislike it when people around me get emotional.	LR/FG
I am used to being emotionally numb.	FG
I do not know how to support people when they become emotional.	FG
I always choose reason over emotion.	FG
Being around people who are upset makes me feel uncomfortable.	FG
I hardly ever experience extreme emotions.	LR/FG
*I am an open book to other people.	LR/FG
I often push through difficult situations without sharing it with anyone.	LR/FG
I don't like to reveal my vulnerability.	LR/FG
*I am very open about my emotions.	LR/FG
*My emotions are a very important part of who I am.	LR
I rarely complain about being stressed or hurt.	LR/FG

I rarely lash out.	LR/FG
If I ever lash out, it tends to be quite explosive.	LR/FG
I often feel numb in situations in which others tend to feel intense emotions.	LR/FG
No matter what I feel on the inside, I make sure I seem fine on the outside.	LR/FG
If I ever lash out, it's only around people that I know very well.	LR/FG
Thinking about crying makes me feel uneasy.	FG
My facial expressions don't always match how I really feel (for example, I laugh when I feel awkward, or I smile when I'm sad).	LR/FG
I consciously put on facial expressions that I think are appropriate in each situation.	LR/FG
Admitting that I feel sad is a sign of weakness.	LR/FG
I tend to bottle up my feelings.	LR
I tell others that I'm fine, even if I'm not.	LR/FG
When I am angry, I become quiet and withdrawn.	LR/FG
I can be passive-aggressive at times.	FG
*I experience positive emotions strongly.	LR
I do not like to dwell on my emotions.	FG
If somebody needs my help, they should communicate it clearly.	FG
People cannot expect me to know how they feel if they don't speak up about it.	FG
I can easily shake off my emotions.	LR/FG
It is difficult for me to stop worrying.	LR
It is difficult for me to control anxious thoughts.	LR
I do not get angry.	LR/FG
To support somebody is to help them find a solution to the problem.	LR/FG
*To support somebody is to help them understand their emotions on a deeper level.	LR/FG
I would generally describe my day-to-day mood as stable.	LR
I rarely laugh out loud.	LR
I take a lot of pride in making cynical, clever jokes.	LR
*People have told me that my face is very expressive.	LR
*When something good happens, I can't wait to share the good news with everyone.	LR/FG
*When something bad happens, talking to other people about it helps me a lot.	LR/FG
People have told me that I'm difficult to read.	LR

Note. Reverse-scored items are indicated with an asterisk (*).

3.7 Discussion

There is a need for a new psychometric measure that allows clinicians and researchers to identify individuals characterised by maladaptive overcontrol. Developing a new self-report measure is the overarching aim of this thesis. The primary goal of this stage of the project was to (1) establish a conceptual framework of maladaptive overcontrol and (2) the functional properties of the measure, and (3) develop an initial item pool for a new self-report measure of the construct.

A theory-driven conceptual framework was established based on the Neurobiosocial Theory for Disorders of Overcontrol (Lynch, 2018), with the four core deficits of maladaptive overcontrol proposed used to operationalise the construct of maladaptive overcontrol and establish its scope. It was hoped that defining maladaptive overcontrol in the context of the four deficits would allow for assessment of both the extent of issues of maladaptive overcontrol as a multi-dimensional construct and the extent of issues pertaining to individual deficits, as well as allow for more

comprehensive and meticulous explanations when using the measure for research purposes.

The initial item pool was developed by the researcher through a combination of deductive and inductive methods for item generation, following best practice recommendations for measure development (Boateng et al., 2018). The deductive stage involved a thorough review of the Neurobiosocial Theory for Disorders of overcontrol and RO DBT literature, in accordance with the theory-driven approach, as well as a review of wider literature on self-control. For the inductive stage, the researcher decided to involve a sample of participants from the general population rather than experts in the field at the inductive stage, so that the items would directly represent individuals' life experiences in relation to the theoretical framework. This is because expert evaluation of the items was planned at a later stage (see Chapter 4). Focus groups were chosen over individual interviews for the purpose of collecting qualitative data regarding the latent construct. A group discussion allows for the participants to directly compare and discuss their experiences regarding the topic of interest, whether differing or similar, with the interaction being the source of data collection – and therefore generating insightful data that would not be accessible if individual interviews were conducted (Morgan, 1996). Both stages of the study were conducted online. The online format facilitated sample beyond a single geographic location, lowering the risk of location bias affecting the results. Further, eliminating the need for participants to travel to an in-person focus group made the process more cost- and time-efficient, suitably for the resource-limited nature of the current programme of work.

In the inductive stage, a pilot focus group study was first conducted with participants with both under- and overcontrolled tendencies. While data generated during the pilot was assessed to be rich and included in the analyses alongside the other discussions, several changes to the study were made following the pilot study (see the Method section for details). Overall, most changes applied following the pilot focus group worked as anticipated. A difficulty was that several participants who completed the screening questionnaire and qualified for the focus group either did not return a consent form either at all, effectively withdrawing from the study, or returned a partially signed consent form, and when asked to fill in the missing sections, did not respond to the e-mail. This resulted in a smaller number of individuals participating in

the focus groups than was initially anticipated, with only two to three participants in each group. Nevertheless, the researcher felt that all domains of the conceptual framework were covered in sufficient depth across the five groups. Of note, some of the participants were considerably less involved in the discussion than others. Nevertheless, the researcher feels that appropriate steps were taken to get every participant involved in the discussion without making them feel targeted.

As a result of the two-stage item generation study, a total of 190 items were generated – approximately three to four times the number of items intended for the final version of the measure. Therefore, the number of items was in line with the recommendation to generate at least double the number of intended items in the first version of the scale (Boateng et al., 2018).

The main challenge of the study was identifying an appropriate sample of participants with a high level of overcontrolled tendencies for the inductive stage. The ASC-WP questionnaire was used to screen potential focus group participants for under- and overcontrolled tendencies, as the only non-clinical scale available designed to screen for under- and overcontrolled tendencies in the context of Lynch's (2018) theoretical model. As such, despite a lack of validity and reliability evidence available, the researcher felt that ASC-WP was the best available scale for the purpose of identifying a relevant sample for the study. However, the data revealed that a disproportionate number of the screened participants screened leaned towards overcontrol, indicating that the ASC-WP may be biased towards overcontrol. To increase the likelihood that the focus group sample indeed exhibited a high level of overcontrolled tendencies, only participants with scores above the mean for the group were invited to participate in focus group discussions. This appeared to have worked well, as during the qualitative stage of the study, the researcher observed that the information shared by most participants was indeed consistent with high overcontrolled tendencies. Nevertheless, the researcher acknowledges a lack of validity evidence for the ASC-WP as a limitation of the current study and recommends that the results are interpreted with caution.

Of note, domain item pools included varying number of positively and negatively worded items across domains. Including a balance of positively and negatively worded items is suggested by some researchers to reduce response biases,

such as acquiescence bias (absent-minded affirmative responding) and agreement bias (avoidance of extreme response options; DeVellis & Thorpe, 2021, Nunnally, 1978). However, others advise against using negatively worded items due to, for instance, more cognitive effort required to understand them and their tendency to form separate dimensions (Salazar, 2015). With mixed evidence as to which approach was best, a decision was made to write each item in a way that appeared to best reflect tendencies towards maladaptive overcontrol based on both the literature review and the focus group discussions, regardless of positive or negative wording and reverse-scoring. It was anticipated that statistical analyses planned for later stages of the project would ultimately indicate which items best assess maladaptive overcontrol.

4. Content validation via expert judgement method

4.1 Introduction

A vital assumption of the measure development process is that the instrument measures what it intends to measure (Beck, 2020). Validity and reliability testing is what allows the scale developer to assess whether the scale appropriately measures the intended construct. There are many different types of validity and reliability evidence that can help assess the properties of the scale, and they are assessed at different stages of the scale development and evaluation process (Murphy & Davidshofer, 2005). According to Boateng and colleagues (2018), the first type of validity to be assessed during the scale development process should be content validity, which is the focus of the current chapter.

Content validity is a type of validity concerned with the degree to which the contents of a new scale are theoretically representative of the construct of interest and reflect a sufficiently wide range of characteristics and behaviours related to that construct without the scale being overly detailed or long (Colton & Covert, 2007; Fernández-Gómez et al., 2020). Ensuring that the item pool accurately represents the theoretical construct that is intended to be measured is a vital step in the early stages of scale development, as it sets the foundation for the later stages (Almanasreh et al., 2019; Boateng et al., 2018). It is typically evaluated based on relevant individuals' opinion on the items in the pool, such as individuals from the target population or, most commonly, individuals with expert knowledge of the area of interest (Boateng et al., 2018; Morgado et al., 2017). Utilising content validity analyses is especially important when scales are being developed for novel and under-researched constructs, as work still needs to be done in empirically mapping out these constructs.

The current chapter presents a study in which carefully selected experts systematically evaluated the OAQ item pool to ensure the adequacy of the scale's contents for measuring the construct of maladaptive overcontrol. Methodological considerations around the expert judgement method are first considered.

4.1.1 The expert judges

The expert judgement method is the most used technique in content validation studies (Boateng et al., 2018). Expert judges are individuals with high levels of expertise regarding the construct of interest (Boateng et al., 2018) developed through either academic and/or non-academic (e.g., clinical) work experience (Fernández-Gómez et al., 2020). Typically, the experts are asked to, based on their professional expertise and experience, assess the degree to which the items in the pool are relevant and reflective of the latent construct (Boateng et al., 2018). It has been recommended that 5-10 expert judges should be involved in content validation studies (Almanasreh et al., 2019). Often, scale developers choose not to identify participating experts by name, in line with the generally accepted ethos that research participants should remain anonymous to the reader. However, Beck (2020) argued that in the case of content validity studies, the experts' identity should – with consent – be communicated in the outputs of the project, for the purpose of shared responsibility for the outcome, as well appropriate recognition for the contribution made to the process. This approach considers experts to be contributors rather than participants.

4.1.2 Approaches for establishing content validity via expert judgement

A variety of approaches for content validation via expert judgement have been described in the literature. One of them is the Delphi method – a multi-round, iterative communication technique designed to allow a panel of experts to achieve consensus about the utility of items in the item pool (Koller et al., 2017). Many variations of the Delphi method exist. The communication between experts can be facilitated face-to-face, online, or by means of physical or electronic written methods (i.e., by post or e-mail), and qualitative, quantitative, or mixed methods can be used. Typically, the experts individually rate items on specified criteria, and the feedback from the entire panel of experts is shared with them, so that answers and feedback can be revised until consensus is achieved (Koller et al., 2017). Grant and Kinney (1992) reported that usually, three rounds of communication are sufficient for the experts to achieve an agreement. While the qualitative Delphi method is likely to generate in-depth and robust data regarding the contents of the scale, due to its multi-round nature, it is also time- and resource-intensive. As such, it has been determined unsuitable for the current programme of research because of the demand on contributors.

A common alternative to the Delphi method is use of a quantitative approach in which experts rate the items on pre-selected criteria using specified numerical values (Koller et al., 2017). These methods allow the scale developer to calculate content validity of the items, the scale domains, and/or the entire scale using statistical estimates (Koller et al., 2017), with the most widely used estimates being the content validity ratio (CVR) and the content validity index (CVI).

The CVR statistic is designed to make decisions about item retention and rejection (Almanasreh et al., 2019). It was first described by Lawshe (1975) and measures the level of agreement between the experts as to how essential a particular item is, with the mean CVR of all the retained items indicating the overall content validity of the scale. The item is only retained in the item pool if it is considered essential by a critical number of experts, based on the table provided by Lawshe (1975). While the CVR is easy to compute, the creator did not describe the original methods of calculating the critical value, making it unclear how to interpret the statistic (Almanasreh et al., 2019).

An alternative content validity statistic is the CVI, typically measures on a 4-point scale, where ratings of 1 or 2 are considered content invalid, and ratings of 3 or 4 are considered content valid (Almanasreh et al., 2019). Almanasreh and colleagues (2019) described and compared three types of CVI – item-level CVI (I-CVI), universal CVI (UA-CVI), and average CVI (Ave-CVI). The I-CVI is computed for each item by dividing the number of experts that rated the item as 3 or 4 on a 4-point scale and dividing it by the total number of experts. Halek and colleagues (2017) reported that with five or more expert judges, the acceptable standard for I-CVI is ≥ 0.78 . Ave-CVI indicates the content validity of the entire instrument. The easiest way to compute Ave-CVI is to sum the I-CVIs of all the items and divide the value by the total number of items. The UA-CVI also represents the content validity of the entire instrument. However, it is more conservative compared to Ave-CVI, as it requires a hundred percent agreement among the experts. It is computed by calculating the proportion of all the items that were rated as 3 or 4 on a 4-point scale by all experts. Acceptable standard for both Ave-CVI and UA-CVI is ≥ 0.8 , with values of ≥ 0.9 indicating excellent content validity (Halek et al., 2017). While it has been recommended to calculate and include both UA-CVI and Ave-CVI for information purposes, the less

conservative Ave-CVI has been indicated to be a preferable statistic for interpretation of instrument's content validity, especially in studies with several expert judges (Almanasreh et al., 2019; Polit & Beck, 2006).

Overall, Almanasreh and colleagues (2019) recommended choosing the CVI method over the CVR method, because it is as simple to compute but also easy to interpret and understand. While the described quantitative approaches are less resource- and time-intensive compared to the Delphi method, there is little room for experts to provide any specific feedback about the items that could guide the modification of the item pool. As such, some researchers advocated for use of approaches that combine the advantages of the described approaches (e.g., Newman et al., 2013). However, despite the widespread agreement on the crucial importance establishing content validity in the scale development process, and the fact that several approaches have been in existence for quite some time, there is no consensus as to a standardised, best-practice approach (Almanasreh et al., 2019; Koller et al., 2017).

4.1.3 Methodological considerations

Regardless of what approach for establishing content validity is used, there are several parameters that ought to be considered when designing an expert judgement content validation study to ensure methodological precision.

Firstly, it has been stressed that expert ratings should be guided by the theoretical framework underpinning the latent construct. Almanasreh et al. (2019) indicated that presenting the conceptual framework that includes the conceptual definition of the latent construct and the operational definitions of identified domains to the experts before the content validation process begins. This ensures that there is a shared understanding between the experts and the scale developer of what the scale is designed to measure.

Secondly, the rating instructions and the criteria based on which the items are meant to be considered must be appropriately presented. In content validity studies, expert judges are typically asked to consider whether the items are, for instance, relevant, meaningful, clear, and complete (Halek et al., 2017). However, according to Beck (2020), simply providing colloquial names for the criteria is not sufficient, as it is impossible to ensure that a particular term has the same meaning to all the judges.

Moreover, Beck suggested that the lack of specific definitions for the terms often communicates the scale developer's insufficient understanding of what exactly they want the experts to judge. He stressed that the criteria should be carefully chosen, clearly defined, and communicated to the experts alongside specific guidance of what is expected of them, with the responses appropriately recorded. Additionally, several researchers highlighted that if quantitative methods are implemented, the experts should also be presented with carefully developed scoring instructions that include indicators of what different numerical values represent (e.g., Almanasreh et al., 2019; Fernández-Gómez et al., 2020).

An example of a carefully developed quantitative rating sheet is one from a content validation study by Fernández-Gómez et al. (2020), which involved a careful translation and adaptation of a previously validated template created by Escobar-Pérez and Cuervo-Martínez (2008). The rating sheet presented clear definitions of four vital criteria – Sufficiency, Clarity, Coherence, and Relevance, as well as provided indicators for the meaning of each numerical value of the rating scale (e.g., for Relevance, 1 – “The removal of the item would not affect the measurement of the dimension”, and 4 – “The item is very relevant and should be included”; p. 5).

4.1.4 The current study

The overarching goal of the current content validity study was to establish whether the newly developed item pool accurately and sufficiently reflected the current theoretical understanding of the construct of maladaptive overcontrol. Typically, content validation studies are focused particularly on item relevance (see e.g., Lawshe, 1975) – or, in other words, on whether the item is essential in the context of the latent construct and must be included in the pool (Fernández-Gómez et al., 2020). For the current study, the researcher decided to include additional relevant rating criteria aside relevance: coherence, clarity, and sufficiency. As such, the aims of the study were to:

1. establish the relevance of the items to the conceptual framework developed based on the Neurobiosocial Theory for Disorders of Overcontrol (Lynch, 2018);
2. establish how coherently the items fit the domains they were placed in during the item generation stage;

3. ensure semantic clarity of the items; and
4. ensure that the domains sufficiently represented the extent of the four deficits of maladaptive overcontrol described by Lynch (2018).

As means of achieving these aims, the researcher decided to combine the advantages of quantitative and qualitative expert feedback and develop a mixed methods rating sheet. The intention was to combine numerical item ratings for the criteria of interest with an opportunity for experts to provide qualitative feedback about the items. The study closely followed methodological recommendations previously discussed in this chapter.

4.2 Method

The study plan was submitted for review by the Schools of Business, Law and Social Sciences Research Ethics Committee (reference ID: 1536443). The application included documents such as participant information sheet, consent forms, and debrief forms – all of which were developed in accordance with the British Psychological Society guidelines and internal Nottingham Trent University guidelines. A favourable ethical opinion was granted for the conduct of the study.

4.2.1 Design

The aim of the study was to establish content validity of a new instrument designed to assess maladaptive overcontrol. A two-round, mixed methods study was conducted to achieve this aim:

First review round.

Expert judges were asked to rate each of the 190 items from the initial item pool ($N_{LFC} = 56$, $N_{LRO} = 41$, $N_{LSC} = 42$, $N_{PIE} = 51$) on three criteria: Relevance, Coherence, and Clarity. After rating all the items within a single domain, the experts were asked to rate that domain on Sufficiency. Lastly, the experts were also given a chance to provide any qualitative comments and suggestions regarding item rejection and modification.

Follow-up review round.

Initially, only one round of ratings was planned due to the project and the experts being time- and resource-limited. However, because substantial changes were made to the

item pool, the researcher decided to conduct a second round of ratings with the same experts, as recommended by Lynn (1986). While she advocated for the entire new version of the item pool to be reviewed in the second round, for the current study, a decision was made to only include the refined ($N = 21$) and added ($N = 4$) items. This was to lower the demand on the expert contributors, and thereby increase the likelihood that they would find the time to provide additional ratings.

As such, experts who completed the rating sheet during the initial stage were asked to rate each item that was either modified or added to the pool following the initial review. Only the three item-level criteria (Relevance, Coherence, and Clarity) were considered due to only a part of the item pool being presented. Quantitative-only data were collected during this round.

4.2.2 Participants

Initial review.

Expert judges were defined as individuals who have extensive knowledge of the Neurobiosocial Theory for Disorders of Overcontrol and RO DBT (Lynch, 2018), due to the measure being theoretically driven. It was also considered important to utilise expertise not only from individuals who have conducted and published research on the topic of overcontrol and RO DBT, but also individuals who have had clinical experience in treating overcontrolled clients. With the model being relatively new and not yet widely empirically tested, expert clinical experience was necessary to ensure that the item pool also reflects how maladaptive overcontrol manifests within clinical populations. As such, to be considered for the study, the individual:

- 1) must have been familiar with the Neurobiosocial Theory for Disorders of Overcontrol and RO DBT (Lynch, 2018); and
- 2) must have either previously conducted research around maladaptive overcontrol and/or radical openness, or involved in treating clients with mental health problems posited to be directly linked to maladaptive overcontrol.

The researcher decided to aim for a minimum of five expert judges to participate in the study, as recommended by Almanasreh et al. (2019). Purposive and snowballing sampling methods were utilised. Fourteen experts from the professional circles of the

researcher and the researcher's supervisory team were approached via e-mail. Two of them provided recommendations for other experts suitable for the study, and as such, seven more experts were invited to take part via e-mail.

Seven of the 21 contacted experts took part in the study. Table 4.1 presents the areas of expertise, and relevant clinical and/or academic experience of the experts involved in the study.

Table 4.1: Experience and expertise of the experts involved in the study.

No.	Experience and expertise
1	Licensed psychologist and psychotherapist with extensive expertise in eating disorders, a co-author of several publications aimed at increasing the understanding of overcontrol in anorexia nervosa.
2	Licensed psychologist, senior RO DBT clinician with expertise in eating disorders, anxiety, mood disorders, and adolescent psychopathology.
3	Licensed psychologist with expertise in depression, chronic anxiety, and anorexia nervosa, senior RO DBT clinician and trainer, and a member of RO DBT research committee.
4	Licensed counsellor with expertise in eating disorders, anxiety, and depression, senior RO DBT clinician, trainer, and supervisor, and a co-author of a forthcoming RO DBT workbook.
5	Licensed clinical psychologist and an expert on personality assessment and personality pathology, who has previously been involved in the development of the theoretical model and assessment of overcontrol in advisory capacity.
6	PhD Psychology holder, co-director of RO DBT Online B.V. and Radically Open Ltd, Trial Manager for multi-site RO DBT randomised control trials, involved in several projects and co-author of several publications involving overcontrol and RO DBT.
7	Licensed psychologist and a certified coach, internationally recognised RO DBT supervisor, and a co-author of several (published and forthcoming) books on RO DBT.

Six of the seven experts returned a completed review sheet. For the experts who returned completed review sheets, both quantitative ratings and qualitative feedback were included in the analyses. The seventh expert indicated a difficulty to rate the items using the rating instructions provided due to the indicators not capturing their thoughts regarding many of the items. The expert returned only a partly completed sheet, and as such, their ratings were not included in the quantitative analysis. However, the expert provided valuable feedback on certain items, which was deemed applicable to the entire scale. Therefore, their qualitative comments were considered alongside those provided by the other six experts.

Follow-up review.

The six experts who had returned completed rating sheets in the initial stage were invited to take part in the follow-up review; five of these experts consented to take part and returned completed follow-up rating sheets.

4.2.3 Materials

Initial review.

The rating sheet utilised in the study was developed via adapting the framework for content validation via expert judgement published by Fernández-Gómez and colleagues (2020). Three item-level and one domain-level criteria were included, with each criterion was explicitly defined, as recommended by Beck (2020). Each of the four criteria was to be rated on the scale from 1 (indicating item/subscale performing poorly on the following criterion) to 4 (indicating item/subscale performing well on the criterion). The meaning of numerical indicators was communicated on the rating sheet. For criteria definitions and rating instructions, see Table 4.2. The researcher also decided to include the developed conceptual framework within the rating sheet alongside the item pool and the rating instructions, as recommended by Almanasreh et al. (2019). Additionally, a *Comments* column was included so that the experts could make qualitative feedback on the items and scales.

Table 4.2: Definitions and indicators for the expert rating criteria.

Criterion and definition	Rating indicators
Relevance (item-level) - considers whether the item is essential or important, i.e., it must be included	1 – removal of the item would not affect the measurement of the domain 2 – the item is somewhat relevant, but another item may already be covering what this item is measuring 3 – the item is rather important 4 – the item is very relevant and should be included
Coherence (item-level) – considers whether the item is logically related to the domain it is measuring and reflects the domain well	1 – the item has no logical relationship to the domain that it is measuring 2 – the item has somewhat logical relationship to the domain it is measuring 3 – the item has moderately logical relationship to the domain it is measuring 4 – the item is logically related to the domain and reflects it very well
Clarity (item-level) - considers whether the item can be understood easily, i.e., the syntax and semantics are appropriate.	1 – the item is completely unclear and should be rewritten 2 – the item requires significant modifications to be easily understood 3 – the item requires some modifications to be easily understood 4 – the item is clear and well-written, with appropriate semantics and syntax

Sufficiency (domain-level) – considers whether the item pool pertaining to the deficit suffices to measure this deficit.	1 – the items are insufficient 2 – a few items must be added to fully assess the deficit 3 – the item measures most aspects of the deficit, but specific items must be added to assess the full extent 4 – the item pool suffices to fully assess the dimension
---	--

Note. Rating framework adapted from Fernández-Gómez et al. (2020).

Follow-up review.

The original rating sheet was adjusted for the purpose of the follow-up review, with the same criterion definitions, rating scale, and numerical indicators provided, and the conceptual framework attached. However, the domain-level Sufficiency criterion was not included within the follow-up round, and neither was the column for qualitative comments. This is because, due to time-limited nature of the PhD, only one follow-up round of expert review was planned, and therefore no more item additions and modifications were anticipated.

4.2.4 Procedure

Initial review.

Potential experts were sent an invitation e-mail with a short introduction to the study, and a participant information sheet and a consent form attached. If wishing to contribute to the study, they were asked to return a signed consent form to the researcher via e-mail. As part of the consent form, experts were also asked to confirm that they meet inclusion criteria for the study, describe their expertise in the area, and sign a declaration on whether they wish to be mentioned in the acknowledgement section in the outputs of the study or prefer to remain anonymous. All experts who took part in the study indicated that they wished to be named as contributors.

After the researcher had received a signed consent form from an expert, an e-mail containing an expert review sheet was sent to them. Experts were instructed to rate each item on the three item-level criteria and each domain on the single domain-level criterion on a scale from one to four. The experts were also encouraged to provide any feedback they may have about the items and domains within the Comments section, with guidance on what type of comments and recommendations the scale developer is looking for within each criterion included.

Experts were asked to return the review sheet to the researcher via e-mail by the deadline agreed at the consent stage. Following return of the completed documents, each expert was sent a debrief sheet and thanked for their contribution to the study.

Follow-up review.

After the decision to conduct a follow-up review has been made, experts who returned completed review sheets in the initial stage were invited to take part in the follow-up study via e-mail. Within the e-mail, the purpose of the follow-up review was explained, and the experts were asked to familiarise themselves with the attached participant information sheet and the consent form provided. The follow-up review sheet was also attached. If the experts wished to take part in the follow-up review, they were instructed to rate each of the 25 follow-up items on each of the three item-level criteria using the same rating scale as in the initial review and asked to return the completed sheet to the researcher alongside the signed consent form by the deadline indicated within the e-mail.

4.3 Data collation and results

4.3.1 Initial review

Firstly, CVI for the Relevance criterion was calculated for each item. Out of the 190 items presented to the experts, 98 items were rejected due to the I-CVI for the Relevance criterion being under 0.78 ($N_{LFC} = 35$, $N_{LRO} = 22$, $N_{LSC} = 16$, $N_{PIE} = 25$), in accordance with the cut-off score recommended in the literature (Halek et al., 2017; see Appendix D for the rejected items).

Secondly, CVIs for Clarity and Coherence were calculated for each of the remaining 92 items. The analysis revealed that three items had a CVI for Coherence of under 0.78 (two in LFC domain and one in LRO domain), and one item had a CVI for Clarity of under 0.78 (LFC domain). Decisions the modification and rejection of these items were based on qualitative feedback provided by the expert judges and resulted in a modification of three items and rejection of one item (see Table 4.3).

Table 4.3: Changes made to items that did not meet the CVI cut-off for the Coherence or Clarity criteria.

Original domain	Original item	I-CVI Coherence	I-CVI Clarity	Applied changes
LFC	I would only break a rule if I knew it was the right thing to do.	1.00	.67	Wording changed to: "I would only break a rule that I do not respect."
LFC	I notice small changes that other people may not see.	.67	1.00	Item replaced with: "When I notice small discrepancies, I must fix them immediately." due to previous lack of items considering sensitivity to discrepancies and overcontrolled individuals' posited tendency to immediately act upon identified discrepancies. Domain changed to LRO. Dislike of change has been covered by other items within the LFC domain.
LFC	I would only ever volunteer to answer a question if I was certain I knew the correct answer.	.67	1.00	Domain changed to LRO.
LRO	Even when I get praised for my work, I still think I could have done it better.	.67	1	No qualitative feedback provided by experts. Item rejected.

Next, CVI was calculated for the Sufficiency criterion for each of the four domains. The analysis revealed that only the LRO domain did not meet the cut-off of 0.78 (LFC = 1, LRO = 0.75, LSC = 1, PIE = 1). Expert feedback indicated that items pertaining to discounting feedback and suspiciousness needed to be added to the domain item pool, and example questions were provided. As such, the following four items were added to the LRO item pool:

- When someone gives me negative feedback, I just ignore it.
- When someone tells me something I do not want to hear about myself, I explain to them why they are wrong about me.
- When someone gives me a compliment, I ask myself what it is that they want from me.
- If people give me positive feedback, it is because they want to manipulate me.

Lastly, the researcher reviewed qualitative feedback to scan for any additional potential issues identified by the experts. Based on expert feedback, semantic changes

were made to seven items, with the items either simplified (N = 4) or the grammar corrected (N = 3; see Table 4.4).

Table 4.4: Items to which semantic changes were made based on expert feedback.

Original domain	Original item	Applied changes
LFC	I like to be prepared for any possibility so that nothing can go wrong.	Simplification. “I like to always be prepared for any possibility.”
LFC	I must be fair, even if I don’t want to at times.	Correction of grammar. “I must be fair, even if I don’t want to be at times.”
LFC	When the stakes are high, I rehearse what I want to say over and over again.	Simplification. “I rehearse what I want to say over and over again.”
LFC	I pride myself in always appearing to be in complete control.	Simplification. “I always try to appear like I am in complete control.”
LFC	Responsibilities should always come first; having fun can wait.	Simplification. “One should always deal with responsibilities before having fun.”
LSC	In social situations, I prefer when people have clearly assigned roles and stick to them.	Correction of grammar. “In social situations, I prefer it when people have clearly assigned roles and stick to them.”
LSC	I am more awkward in social situations than most people.	Correction of grammar. “I behave more awkwardly in social situations than most people do.”

One expert noted that while the LSC domain should focus on relationships with other people, including empathic perception, the PIE domain should focus on one’s internal experiences of emotions. In accordance with expert feedback, modifications were made to four items (see Table 4.5).

Table 4.5: Items modified based on expert feedback indicating a better fit to a different domain based on domain definition.

Original domain	Original item	Applied changes
LSC	*I am a warm and affectionate person.	Domain changed to PIE.
PIE	I do not know how to support people when they become emotional.	Domain changed to LSC. Wording changed to: “When other people become upset, I often don’t know how to support them.”

PIE	Being around people who are upset makes me feel uncomfortable.	Domain changed to LSC. Wording changed to: “I dislike being around people who are upset.”
PIE	I strongly dislike it when people around me get emotional.	High likelihood of for cross-correlation between LSC and PIE. Item replaced with another item that was suggested by an expert: “When I see another person crying, I find it difficult to understand why they would do this in public.” and moved to LSC domain.

Note. Reverse-scored items are indicated with an asterisk (*).

One expert also expressed concerns regarding several items in the item pool that used words such as “upset” in the context of the respondent’s own emotional experience. The expert indicated that such words increase correlation with constructs such as negative affectivity and neuroticism. The researcher noted that this could potentially introduce a correlation across domains, especially with the PIE domain’s focus being emotionality and affect, which would considerably complicate the structure of the measure. As such, all the items that used such wording but were not a part of PIE domain were modified (see Table 4.6).

Following these analyses, 99 items were rejected. All the items that were modified (N = 21, including one item replaced by a new item) and added (N = 4) were submitted to experts for a follow-up review.

Table 4.6: Items to which changes were applied due to potential correlation with constructs such as emotionality and affect.

Domain	Original item	Modified item
LFC	Small changes to my daily routine upset me.	I dislike it when my daily routine changes even slightly.
LFC	Lack of structure and order upsets me.	Structure and order are very important to me.
LFC	I persist at tasks even if they cause me distress.	I have the ability to persist at difficult tasks for long periods of time.
LSC	I feel that I am an outsider.	I am an outsider.
LSC	I feel different from other people.	I am different from other people.
LSC	I feel distant from other people.	I am distant from other people.
LSC	I often feel misunderstood by other people.	I am often misunderstood by other people.

4.3.2 Follow-up review

For each item included in the follow-up review, a CVI was calculated for Relevance, Coherence, and Clarity. Out of the 25 follow-up items, 20 items had a CVI of above 0.78 for all three criteria, and as such, new versions of these items were retained. The remaining five items met the cut-off for Coherence and Clarity criteria, but not for Relevance. Each of these items was then reviewed individually, and decisions were made regarding retention and rejection (see Table 4.7).

Table 4.7: Decisions regarding retention and rejection of the follow-up items that did not meet the CVI cut-off for the Relevance criterion.

Item	Domain	I-CVI Relevance	Decision
I would only break a rule that I do not respect.	LFC	.60	Previous version of the item met the cut-off for Relevance (.833), but not for Clarity (.667). Item rejected.
I always try to appear like I am in complete control.	LFC	.60	Previous version of the item (<i>I pride myself in always appearing to be in complete control.</i>) with CVI above .780 for all three criteria was retained.
When I see another person crying, I find it difficult to understand why they would do this in public.	LSC	.60	Item does not have a previous version and was only included in the follow-up review. Item was retained in the item pool despite not meeting the cut-off due to lack of other items covering this aspect of the construct.
I am an outsider.	LSC	.60	Previous version of the item (<i>I feel that I am an outsider.</i>) with CVI above .780 for all three criteria was retained.
*I am a warm and affectionate person.	PIE	.75	The item was previously listed in the LSC domain (same wording). With only four experts having rated the item in the follow-up review, and expert feedback previously indicating that the item should be included in PIE domain due to relating to self rather than others, the item was retained in the PIE domain despite not meeting the cut-off.

Note. Reverse-scored items are indicated with an asterisk (*).

4.3.3 The new item pool

Following the two rounds of review by expert judges, a new version of the item pool was formed, with a total of 94 items across the four domains ($N_{LFC} = 18$, $N_{LRO} = 24$, $N_{LSC} = 28$, $N_{PIE} = 24$; see Table 4.8).

Table 4.8: Refined item pool following two rounds of expert review.

Domain	Item
LFC	I like to always be prepared for any possibility.
LFC	I like to make detailed plans for everything.
LFC	I feel the need to fix all problems immediately.
LFC	I have strong opinions about how things should be done.
LFC	I strongly dislike it when plans change unexpectedly.
LFC	I dislike it when my daily routine changes even slightly.
LFC	Structure and order are very important to me.
LFC	I must be fair, even if I don't want to be at times.
LFC	I rehearse what I want to say over and over again.
LFC	I persist at tasks even if they cause me distress.
LFC	Being correct is more important to me than it is to most people.
LFC	I like my life well-structured and predictable.
LFC	It's important to me that people adhere to social norms and standards.
LFC	Having self-control is more important to me than it is to most people.
LFC	I must always do what I believe is right.
LFC	I pride myself in always appearing to be in complete control.
LFC	I cannot stop thinking about a problem until I find a solution.
LFC	One should always deal with responsibilities before having fun.
LRO	I would only ever volunteer to answer a question if I was certain I knew the correct answer.
LRO	When I notice small discrepancies, I must fix them immediately.
LRO	When someone gives me negative feedback, I just ignore it.
LRO	I carefully consider potential consequences before taking a risk.
LRO	*I welcome critical feedback even if I did not ask for it.
LRO	I purposefully avoid situations in which I could be criticised.
LRO	I purposefully avoid situations in which I could be seen as weak or incompetent.
LRO	*I often do things just for fun.
LRO	*I am always willing to spontaneously try new things.
LRO	I am always on high alert.
LRO	I strongly believe in the saying "better safe than sorry".
LRO	*Unpredictability is exciting.
LRO	The world is a scary place.
LRO	When I make a mistake in front of others, I feel more embarrassed than most people would.
LRO	Avoiding negative consequences is a better motivator than gaining something positive.
LRO	I often see a threat where others may not.
LRO	I dislike uncertainty more than most people.
LRO	*Life without risks is no fun.
LRO	*I am always happy to get out of my comfort zone and try something new.
LRO	I would do anything to avoid being the centre of attention.
LRO	Embarrassing myself in front of other people is my worst fear.
LRO	When someone tells me something I do not want to hear about myself, I explain to them why they are wrong about me.
LRO	When someone gives me a compliment, I ask myself what it is that they want from me.
LRO	If people give me positive feedback, it is because they want to manipulate me.
LSC	In social situations, I prefer when people have clearly assigned roles and stick to them.
LSC	I rarely feel connected to other people.
LSC	I have few close friends, if any.
LSC	I feel that I am an outsider.
LSC	*Being around other people makes me feel alive.
LSC	*I am a very sociable person.
LSC	I generally prefer being on my own rather than around other people.
LSC	I dislike pointless social interactions.
LSC	I often yearn to be alone for a while after participating in social events.
LSC	When disagreements arise, I withdraw from the situation.
LSC	It takes me a long time to warm up to other people.
LSC	I dislike parties.
LSC	*I love meeting new people.
LSC	*I don't mind starting a conversation with a new person.

LSC	I often struggle to understand another person's perspective.
LSC	*Being around other people helps me relax.
LSC	I envy people who do better than me.
LSC	I only ever cry on my own or in front of people that I am very close to.
LSC	My relationships with other people are rather superficial.
LSC	I behave more awkwardly in social situations than most people do.
LSC	I tend to hold grudges for a long time.
LSC	I struggle to believe in real, unconditional love.
LSC	When other people become upset, I often don't know how to support them.
LSC	I dislike being around people who are upset.
LSC	I am different from other people.
LSC	When I see another person crying, I find it difficult to understand why they would do this in public.
LSC	I am distant from other people.
LSC	I am often misunderstood by other people.
PIE	I rarely get very excited.
PIE	I tend to discount and "push down" my emotions.
PIE	Being emotional is a sign of weakness.
PIE	I am used to being emotionally numb.
PIE	I always choose reason over emotion.
PIE	*I am an open book to other people.
PIE	I often push through difficult situations without sharing it with anyone.
PIE	I don't like to reveal my vulnerability.
PIE	I rarely complain about being stressed or hurt.
PIE	I often feel numb in situations in which others tend to feel intense emotions.
PIE	No matter what I feel on the inside, I make sure I seem fine on the outside.
PIE	Thinking about crying makes me feel uneasy.
PIE	My facial expressions don't always match how I really feel (for example, I laugh when I feel awkward, or I smile when I'm sad).
PIE	I consciously put on facial expressions that I think are appropriate in a given situation.
PIE	I tend to bottle up my feelings.
PIE	I tell others that I'm fine, even if I'm not.
PIE	*I experience positive emotions strongly.
PIE	I do not like to dwell on my emotions.
PIE	It is difficult for me to stop worrying.
PIE	To support somebody is to help them find a solution to the problem.
PIE	I rarely laugh out loud.
PIE	*When something bad happens, talking to other people about it helps me a lot.
PIE	People have told me that I'm difficult to read.
PIE	*I am a warm and affectionate person.

Note. Reverse-scored items are indicated with an asterisk (*).

Ave-CVI values were calculated for each of the four domains on each of the three criteria, indicating adequate content validity of each of the domains and the entire item pool. The values indicated acceptable content validity of all domains, and excellent content validity of the entire instrument on all three criteria (Halek et al., 2017). UA-CVI values were also calculated for informative purposes and, unsurprisingly for a study with several expert judges, did not meet the rigorous cut-off except for the Clarity criterion. For Ave-CVI and UA-CVI values, see Table 4.9.

Table 4.9: Domain and combined Ave-CVI for Relevance, Coherence, and Clarity criteria after the follow-up review.

Domain	Ave-CVI			UA-CVI		
	Relevance	Coherence	Clarity	Relevance	Coherence	Clarity
LFC	.93	.95	.99	.61	.72	.94
LRO	.89	.95	.98	.38	.71	.88
LSC	.85	.90	.95	.36	.57	.86
PIE	.89	.94	1.00	.38	.67	1.00
Global	.90	.94	.99	.43	.67	.93

4.4 Discussion

Content validation of a new instrument is a crucial part of the scale development process (Boateng et al., 2018). It aims to ensure that the item pool accurately reflects how the content manifests in the population (Almanasreh et al., 2019). The current study aimed to establish content validity of a newly developed scale designed to assess maladaptive overcontrol using the expert judgement method.

The expert judgement method of content validation relies on the opinion of individuals who have expert knowledge and expertise regarding the latent construct (Boateng et al., 2018), and a correct selection of experts is crucial in content validation studies (Fernández-Gómez et al., 2020). For the current study, experts were defined as professionals familiar with the Neurobiosocial Theory for Disorder of Overcontrol (Lynch, 2018) – a theoretical model underpinning the development of the instrument – who have conducted research on overcontrol or treated clients who were assessed as overcontrolled.

While there is little consensus in literature as to how many experts should take part in a content validation study (Fernández-Gómez et al., 2020), Alamanasreh and colleagues (2019) suggested that 5-10 experts are sufficient. Following this recommendation, the study aimed for a minimum of five expert judges, which has been achieved. However, it is worth noting that only just over a quarter of all the invited experts decided to take part in the study, highlighting the difficulties of conducting content validity studies that involve busy professionals to find time to commit to rating a large pool of items. This was further reflected in one of the experts only sending a partially completed sheet. Nevertheless, due to many qualitative comments, the researcher decided to utilise the qualitative part of the feedback only, which has proven very helpful in refining the item pool. Quantitative ratings that were included were not

considered due to the expert reporting that the given indicators did not reflect their opinion on the items well. While this initially raised the researcher's concerns about the chosen indicators, they have been previously successfully used in content validation study by Fernández-Gómez and colleagues (2020), and no other experts reported such issues in the current study.

Importantly, the experts in the study were given a chance to be recognised in the acknowledgement section in the outputs of this study, in accordance with Beck's (2020) suggestion that experts should be named for the purpose of both recognition of their contribution and shared responsibility for the outcomes of the project. All the experts consented to be named within the outputs, indicating that expert judges may indeed appreciate the opportunity to be recognised for utilising their expertise to help the instrument development process. Nevertheless, the researcher believes that as with any study respondent, the experts should always be given a chance to remain anonymous.

Considering the lack of a gold-standard method for expert judgement content validation (Alamanresh et al., 2019; Koller et al., 2017), one of the most challenging decisions was the choice of method for the study. The researcher decided to utilise a mixed methods rating sheet, with the intention was to use numerical ratings to calculate CVI values for each of the domains and the complete measure, and to use the qualitative comments as guidance when refining the item pool. The rating sheet was carefully developed, and included the conceptual framework, clear rating instructions, specific definitions of the rating criteria, and guidance on how the numerical values should be interpreted. This was to reduce the possibility for ambiguous interpretation of various elements of the study between the experts and make the rating process more straightforward and less time-consuming. The researcher feels that the study design provided a necessary balance between the depth and quality of the data and the time- and resource limitation of the current PhD programme. Nevertheless, the abundance of methods available for this type of study and lack of consensus as to the gold standard approach resulted in the process being complex, effortful, and, especially at the beginning, confusing. The field of psychometrics would certainly benefit from a standardised approach to content validation via expert judgement method.

The study design also allowed for efficient data collection when it had been decided that a second, quantitative-only round was needed. While adding an additional round of ratings extended the study beyond the originally planned timeline, the researcher believes that it was necessary to ensure the robustness of the study. Including only modified and added items rather than a full item pool in the second round of ratings was not an ideal solution but was motivated by experts' concerns about the length of the item pool and the time required to fill in the rating sheet. As such, it allowed the researcher to achieve the balance between the quality of the data and the potential time and resource limitations, as reflected in high number of the experts agreeing to take part in the second round.

Overall, the results of the study indicated sufficient content validity of the newly developed instrument, as per the Ave-CVI values for each of the domains and the composite measure. While, except for the Clarity criterion, the UA-CVI values do not meet the recommended cut-off, they should be interpreted with caution. Acceptable standard for UA-CVI is difficult to achieve in studies with several expert judges (Almanasreh et al., 2019), and therefore the provided values were provided for informative purposes rather than as appropriate indicators of content validity of the instrument. The number of items in the pool was also reduced as a result of this study, as expected at this stage of the psychometric process. Despite satisfactory content validity evidenced in the conducted study, it is recommended that content validity for the measure is further studied in the future among specific populations, with the item pool adjusted for the population needs.

5. Cognitive interviews with the target population

5.1 Introduction

Psychometric literature indicates that development and validation of a new scale should involve not only input from experts in the field, but also the potential future respondents – or, in other words, the target population (Boateng et al., 2018). Aside the target population's involvement in the item development stage (see Chapter 3), their opinion on the items in the pool may also be sought for pre-testing and content validation purposes. The current chapter briefly discusses the aims of involving potential future respondents in these stages of the psychometric process and presents a cognitive interview study with the target population sample.

5.1.1 Pre-testing the scale

The input from potential future respondents is also used in the pre-testing stage of the psychometric process. It involves administering the questionnaire to a sample of participants from the target population to ensure semantic clarity of the items in the pool and consistent, unambiguous comprehension of the items by the intended users (Drennan, 2003). The collected data are used to identify potential sources of confusion that may have not previously been apparent to the individuals involved in the scale development, leading to a subsequent refinement of the problematic questionnaire items (Reynolds et al., 1993). Several issues were previously identified in relation to item clarity by Conrad and Blair (1996):

- (1) lexical problems, representing issues around understanding the meaning of certain words;
- (2) temporal problems, representing ambiguity around the meaning of temporal terms, such as the length of time that should be considered when arriving at a response;
- (3) exclusion/inclusion problems, representing ambiguity regarding whether a certain concept should or should not be included within scope of the item;
- (4) logical problems, such as: a coordinating conjunction being used to connect words that are not univocal to the respondent, creating ambiguity in how to choose a response; questions including incorrect presumptions about the respondents; questions containing contradictions; and

- (5) computational problems, representing issues with the respondents' processing and manipulating information that do not fit any of the previous categories, such as issues with task performance, recall of relevant information, or complex item structure causing overt cognitive load.

Pretesting is considered a crucial step in the scale development process, as it helps to prevent these issues and in turn increase questionnaire response rates and reduce sampling error (Drennan, 2003).

5.1.2 Content validation via target population judgement

Target population's involvement may also be pursued when validating the content of the item pool – either as an addition or an alternative to the expert judgement method (Boateng et al., 2018). This is to ensure that the item pool reflects the true lived experience of individuals in relation to the latent construct. While the target population judgement is used less commonly compared to the expert judgement, some researchers suggested that target population's input should be a standard practice (Clark & Watson, 1995; Morgado et al., 2017). Boateng and colleagues (2018) argued that the intended users of the scale should be perceived as the experts in evaluating face validity of a construct – conceptualised as part of content validity – due to their lived experience. They recommended that, where possible, both expert and target population judgement should be sought as part of the content validation process.

However, utilising the target population judgement in the context of content validation can be challenging in cases where it is difficult to ensure that the recruited sample is indeed characterised by appropriate levels of the latent construct. This may happen if an alternative measure of the latent construct is lacking, or if there is little empirical evidence that would support utilising other instruments that measure concepts linked to the latent construct for sample screening purposes. For instance, in the context of the current programme of research, the new questionnaire is intended to identify maladaptive levels of overcontrol. As such, a highly overcontrolled sample would be required for a target population content validation study – and appropriate screening measure of overcontrolled tendencies is absent, with the only non-clinical screening measure that aligns with the theoretical framework, the ASC-WP (Lynch, 2018), lacking validity and reliability evidence. With the utilised theoretical model

being relatively new, there is also little empirical evidence that would directly support the relationship between maladaptive overcontrol as conceptualised by Lynch (2018) and other psychological concepts that he posits to be linked to maladaptive overcontrol. As a result, the selection of an appropriate sample for the purpose of further validating the content of the item pool could prove challenging.

5.1.3 Cognitive interviewing

Cognitive interviewing has been recommended as a method of getting the target population involved in the process of scale development, both for the purpose of content validation and pre-testing of the scale (Boateng et al., 2018). The technique involves individually administering the measure to a sample of participants during an interview with an aim to collect feedback on the items in real time to identify any issues with the items in the pool that could compromise data quality (Shiyanbola et al., 2019). The recommended sample size for cognitive interview studies varies from around 5 to 30 participants (see e.g., Willis, 2004; Willis & Artino, 2013), with literature indicating that larger sample sizes allow for better detection of issues with the items in the pool (Blair & Conrad, 2011).

During cognitive interviews, relevant information about the items is typically obtained through use of verbal cognitive probing – a technique which involves asking the participants a series of questions that aim to elicit detailed information about their understanding and interpretation of the questionnaire items (Shiyanbola et al., 2019). Willis and Artino (2013) noted that probes may be both proactive – developed ahead of the interviews – and reactive – developed during the interview, in response to the participants’ comments. They also distinguished several types of probes, including:

- (1) general probes that may aim to establish the process of arriving at a particular answer, understand potential hesitation of a participant when attempting to respond to an item, or understand whether choosing a response to a particular item was easy or difficult;
- (2) specific probes that ask about specific aspects of an item or response to an item that are of interest to the scale developer;
- (3) comprehension/interpretation probes that explicitly ask about the meaning that an item or a part of an item has to the participant;

- (4) paraphrasing probes that require the participant to explain an item's meaning in their own words;
- (5) confidence judgement probes that aim to establish how difficult it is for the participant to accurately answer an item; and
- (6) recall probes that require the participant to explain their thought process that led to arriving at a particular answer to an item.

The technique of verbal probing does, however, have drawbacks. Conrad and colleagues (1999) noted that verbal probes – presumably, the more specific types of probes – have the potential to guide the participant's thinking towards aspects they would not have considered if only the think aloud method was used. While some see it as an advantage of the method due to the researcher receiving exactly the answers they are interested in, others see it as a source of bias, decreasing the objectivity of the generated data (Conrad et al., 1999; Gerber & Wellens, 1997).

Aside from probing questions, a 'think aloud' method is also commonly used to generate cognitive interview data (Conrad et al., 1999; Willis & Artino, 2013). In this method, participants are encouraged to freely report their thought process while deciding on the response option, rather than retrospectively asked on how they arrived at an answer. While this has been argued to increase objectivity of participant responses when compared to verbal probing, the approach is limited to the information that the respondent is explicitly aware of and evaluates as important enough to share with the scale developed during the interview – in which case probing questions may prove useful (Conrad et al., 1999).

According to Willis and Artino (2013), analysing the collected data typically involves comparing and interpreting notes on item functioning taken during or after each interview. Such notes may, for example, denote that a participant's understanding of an item was different to what the researcher intended, or that a certain word was reported to be unclear or ambiguous. The researcher would then typically seek convergence across participants and alignment of interpretations with the conceptual framework and refine or remove items that prove perplexing. Modified items may subsequently be tested in an additional round of interviews. However, Conrad and colleagues (1999) noted that scale developers during the analysis stage often overly focus on issues with meaning of the items or words and phrases within the items and

overlook other types of problems. As such, both Conrad and colleagues (1999) and Conrad and Blair (1996) advocated for explicitly labelling the type of problem encountered during the analysis in line with their classification (see section 5.1.1). At the same time, Conrad and colleagues acknowledged that, based on some studies they conducted, many of the problems may prove impossible to code in such manner.

5.1.3 The current study

Despite potential methodological drawbacks, cognitive interviewing technique has been labelled as crucial to the scale development and remains one of the most common approaches of getting the target population involved in the process (Boateng et al., 2018; Willis & Artino, 2013). As such, a qualitative target population study utilising the cognitive interviewing technique was arranged to identify potential issues with the items in the OAQ, with the aim to further refine the item pool based on the collected data. The aims of the study were to:

- (1) pre-test the scale and ensure the consistency of understanding of the items between participants, as well as their semantic simplicity and clarity; and
- (2) further validate the content and ensure that the item pool accurately reflects how the context manifests in the population, this time by utilising the target population judgement method.

The cognitive interviewing technique allows for the scale developer to ensure that participants understand the items in the way they were intended, as well as to understand participants' lived experience in relation to the latent construct. This directly aligns with the objectives of both pre-testing and content validation, making the cognitive interviewing technique a viable choice for this stage of the scale development process (Boateng et al., 2018).

Of note, the distinction between pre-testing and content validation via target population involvement is not consistent across literature. It appears that many researchers use these terms interchangeably or understand content validation to constitute a part of pre-testing when cognitive interviews are used (see e.g., Drennan, 2003; Hilton, 2017; Shiyanbola et al., 2019; Blair & Conrad, 2011). In contrast, Boateng and colleagues (2018) distinguish between the target population content validation and pre-testing. They refer to content validation as means of ensuring that

the questionnaire items accurately reflect the latent construct, and to pre-testing as means of ensuring clarity of language and understanding of the items. To validate the content of a measure designed to identify maladaptive levels of a construct, as is the case in the current thesis, a population high on that construct would be required for the study. It could be argued that this is not necessary when aiming to identify issues with item comprehension or semantic clarity, as these should be consistent regardless of the participants' standing on the latent construct. As such, target population content validation and pre-testing were considered distinct in this thesis.

5.2 Method

The study plan was submitted for review by the Schools of Business, Law and Social Sciences Research Ethics Committee (reference ID: 1536443). The application included documents such as participant information sheet, consent forms, and debrief forms – all of which were developed in accordance with the British Psychological Society guidelines and internal Nottingham Trent University guidelines. A favourable ethical opinion was granted for the conduct of the study.

5.2.1 Design.

The aims of the study were to pre-test the new version of the questionnaire and further validate its content. A two-part, mixed-methods online study was conducted:

Part 1. Participants were initially screened for high levels of overcontrolled tendencies using a self-report ASC-WP questionnaire (Lynch, 2018).

Part 2. Online cognitive interviews were conducted with participants who scored highly on overcontrolled tendencies.

Only one round of testing was conducted due to time and resource limitations.

5.2.2 Participants.

First recruitment round.

A highly overcontrolled sample was required for the purpose of content validation. As such, the target population was initially defined to include any (English-speaking) individual suspected to have high overcontrolled tendencies in the light of the

theoretical model. The researcher decided to screen participants for overcontrolled tendencies using the ASC-WP, which was previously used in the item development study (see Chapter 3). While it previously appeared to be biased towards identifying overcontrolled tendencies, the researcher observed that after applying a high cut-off score, the participants selected for the item development study exhibited characteristics consistent with the tendencies indicated by the results of the screening questionnaire. Purposive sampling was utilised, with the screening survey initially advertised exclusively on social media groups concerned with mental health (e.g., Facebook and Reddit communities regarding eating disorders, personality disorders, depression, anxiety, obsessive-compulsive disorder, autism spectrum disorders, etc.). This was to further increase the likelihood that participants with amplified levels of maladaptive overcontrol take part in the current study. The researcher initially intended to advertise only on groups pertaining to mental health problems that Lynch (2018) linked to maladaptive overcontrol. However, upon consideration of a high level of comorbidity, similarity of symptoms, and arbitrary boundaries between categorical mental health disorders (for a discussion, see e.g., Forbes et al., 2016; Kotov et al., 2017; Krueger & Markon, 2006, 2011; Widiger & Clark, 2000), a decision was made not to limit advertisement of the study to these specific issues.

To take part in the screening survey, the participants needed to be 18 years old or over and speak fluent English. A total of 104 responses to the screening survey were recorded. Responses with no e-mail address provided ($N = 17$, 16.35%) and a completion time of under two minutes ($N = 9$, 8.65%) were removed. A further four responses (3.85%) were deleted due to incomplete data (no answers to the screening questionnaire items were provided).

The remaining 74 responses were complete and included in the analyses. The results revealed that 53 participants (71.62%) leaned towards overcontrol, with an average of 75.59% of responses in the overcontrolled column ($SD = 11.96$, range = 51.06-97.87%)⁵. Every participant whose percentage of responses in the overcontrolled column was above the group mean ($N = 26$, 49.06%) was invited to participate in a

⁵ Of note, for participants who leaned towards undercontrol ($N = 21$, 28.38%), the mean percentage of responses in the undercontrolled column was 39.62% ($SD = 10.54$, range = 10.64-48.94%).

cognitive interview via e-mail. Nineteen invited participants scheduled interviews, of which only 11 interviews took place due to no-shows.

The final sample of interviewees recruited online included seven male and four female participants with a mean age of 27.55 years ($SD = 4.52$, range = 22-39 years old; 8 Black, 2 White, 1 Asian). Nine participants (81.81%) reported United Kingdom as their country of birth and current residence, and two participants (18.18%) indicated that they were born and resided outside of the United Kingdom. Ten participants (90.91%) reported that English was their first language, and one participant (9.09%) reported that it was their preferred language. For more details on the demographics of the sample, please see Appendix E.

Please note that the ASC-WP scores and demographic information provided above should be interpreted with caution. This is because several issues transpired once cognitive interviews began. Firstly, there were indications that screening was inadequate. There were discrepancies between reported and observed demographics, meaning that the answers to the screening questionnaire could have not been meaningful. Secondly, the content of the interviews with several participants suggested that their typical behaviours and characteristics did not fit the description of an overcontrolled individual as per Lynch's (2018) description. It could not be verified that the participants were indeed characterised by overcontrolled tendencies, and as such, the sample was not suitable for testing whether the item pool accurately reflects how overcontrol manifests. As a result, the aim of further validating the content was dropped, and the researcher decided to focus solely on pre-testing. Item understanding should be consistent across participants regardless of the level of the latent trait, because the scale is likely to be administered to respondents with varying levels of the construct in the future. As such, the pre-testing stage did not require the study participants to be highly overcontrolled. The target population definition for the study was therefore adjusted to include any English-speaking individual in the general population, regardless of their tendencies towards under- or overcontrol.

The researcher initially aimed for 24 cognitive interviews to be conducted, which would have resulted in 12 participants per scale domain (two domains per person) and allowed for some no-shows. Due to a high no-show rate, only 11 interviews were conducted. While this met the minimum requirement of five

participants per domain, given the issues described in this section, the researcher deciding to conduct a second round of recruitment to increase the total sample. However, due to issues encountered in the first round where people unknown to the researcher were recruited, a decision was made to opt for a convenience sample of participants known to the researcher.

Second recruitment round.

Convenience sampling was utilised, with the study adverts being sent directly to individuals within the researcher's known networks. To take part, individuals needed to be 18 years old or over and speak fluent English.

A total of 14 responses to the screening survey were recorded during this stage of recruitment. Of those, 12 participants (85.71%) were indicated to have overcontrolled tendencies, with an average of 73.94% of responses in the corresponding column (SD = 11.41, range = 59.57-89.36%).⁶ To ensure consistency of scores across the two rounds of recruitment, all participants with overcontrolled tendencies who scored above the mean used in the first round of recruitment (75.59% of responses in the overcontrolled column) were invited to book cognitive interviews (N = 5, 41.67%).⁷

All invited participants took part in the cognitive interviews. The final sample from this round of recruitment consisted of 3 male and 2 female participants with a mean age of 28.20 years (SD = 8.35, range = 23-43 years old; 4 White, 1 White/Asian). Three participants indicated English to be their first language, and two indicated English to be their preferred language. For more details on the demographics of the sample, please see Appendix F.

5.2.3 Materials

Part 1. The ASC-WP questionnaire was used to screen participants for overcontrolled tendencies (see section 3.5.1 of Chapter 3 for questionnaire

⁶ Of note, for participants who leaned towards undercontrol (N = 2, 14.29%), the mean percentage of responses in the undercontrolled column was 39.36% (SD = 13.54, range = 29.79-48.94%).

⁷ The interview invitations in the second round of recruitment were sent as soon as a qualifying participant had been identified. The recruitment was stopped after a total of 16 interviews across the two rounds of recruitment were conducted due to time limitations of the PhD programme.

description). The study also included a demographic questionnaire with questions regarding the following variables: age, country of origin, country of residence, ethnicity, sex assigned at birth, gender identity, marital status, education level, and a question on whether if English is a participant's first or preferred language.

Part 2. Items from the newly refined pool were presented to participants during the interview (two domains per participant). Cognitive probing questions and the think aloud method were used. Reactive probes were also used where deemed necessary. Example probing questions are presented below:

- “How did you come up with your answer?” (Willis & Artino, 2013),
- “How did you arrive at that answer?” (Willis & Artino, 2013),
- “Tell me more about why you answered [response option] for this question.” (Shiyanbola et al., 2019),
- “I noticed you hesitated. Tell me what you were thinking.” (Willis & Artino, 2013),
- “Tell me more about that.” (Willis & Artino, 2013),
- “What does the word/term [salient word/term within the question] mean to you?” (Willis & Artino, 2013),
- “Can you repeat the question I just asked in your own words?” (Willis & Artino, 2013),
- “Was that easy or hard to answer?” (Willis & Artino, 2013),
- How sure are you about your answer? (adapted from Willis & Artino, 2013)
- You said that [participant's statement]. Why is that? (adapted from Willis & Artino, 2013).

5.2.4 Procedure.

Part 1. Within the Qualtrics screening survey, participants were presented with a participant information sheet explaining the purpose of the study, inclusion criteria, withdrawal rights both during and after the study, and explained how the data would be stored and used. To proceed, the participants were asked to sign a consent form, where they would confirm that they meet the inclusion criteria, that they were happy for the data to be used for research purposes, and that they were happy to be contacted by the researcher if selected for the focus group interview. They were also asked to

provide an e-mail address so that the researcher could contact them if they were selected to participate in the focus group discussion. E-mail was also used as a unique identifier in case of data withdrawal.

After signing the consent form, participants were presented with the demographic questionnaire, with all questions in the section being optional. The participants were presented with the ASC-WP questionnaire to screen for overcontrolled and undercontrolled tendencies. They were instructed to, for each pair, select the word or phrase that they consider to be more characteristic of them, and advised not to overthink the answers and go with their intuition. The participants were also advised that they may use a dictionary in case they are unfamiliar with one of the words or phrases.

After completing the ASC-WP, participants were shown a debrief screen where they were thanked for participation, reminded about the withdrawal rights, and provided with contact details for the researcher, the researcher's supervisory team, and points of support they could contact if the study caused them any distress.

Part 2. Participants selected to take part in the cognitive interviews based on their ASC-WP scores were invited via e-mail. They were asked to sign up for available interview slots using a Doodle poll. Once their chosen slot had been confirmed, they were directed to a Qualtrics form which included a participant information sheet and consent form. The participant information sheet and the consent form included the same information as previously, as well as information about recording of the interviews and handling of the recordings and transcripts. The researcher verified that the consent form had been signed before the interview began.

At the start of the interview, the researcher outlined the purpose of the study and reminded participants about recording of the interview and their withdrawal rights. The researcher explained in detail what the study would involve. Participants were informed that they would see questionnaire items and possible response options on their screen. They were verbally instructed on how to complete the questionnaire. The researcher explained that she would read the items out loud and ask the participant some probing questions regarding their choice of response and understanding of items or phrases. Participants were also encouraged to use the think aloud method if

comfortable and reassured that there were no wrong or right answers to any of the items. They were also reminded that they are free not to answer any questions that they were uncomfortable answering. Each participant was given a chance to ask any questions.

The researcher turned on the recording and the interview began. Where a participant answer to an item was like the answers given by one of the previous participants, the researcher asked iterative questions as means of confirming that the understanding of questions was comparable. After going through all the items within the domains assigned to a participant, the researcher stopped the recording, debriefed the participant, and thanked them for their contribution to the study. A debrief sheet and the incentive voucher were sent to participants via e-mail.

5.3 Data collation and results

After the interview stage had been completed, the researcher listened to the recording of each of the sixteen interviews with an automatically generated transcript at hand and annotated participant feedback about each item in an Excel spreadsheet, with a focus on semantic clarity and participants' understanding of the item. The feedback from all participants was compared, and the item list was refined based on the feedback. Changes were made to 23 items ($N_{LFC} = 4$; $N_{LRO} = 3$; $N_{LSC} = 6$; $N_{PIE} = 10$; see Table 5.1), including two PIE items being split into two. While problems found were not explicitly labelled, the different types of problems identified by Conrad and Blake (1996) were kept in mind when analysing the notes.

Table 5.1: Changes made to the items following cognitive interviews with the potential respondents.

Domain	Old version	New version	Justification
LFC	I rehearse what I want to say over and over again.	In day-to-day situations, I rehearse what I want to say over and over again.	Most participants understood the question in terms of preparation for important meetings or work presentations rather than everyday life rehearsal deemed typical to OC. The phrase "in day-to-day situations" was added.
LFC	It's important to me that people adhere to social norms and standards.	I get annoyed when people don't adhere to social norms and standards.	Similar responses of participants' responses indicated that adherence to social norms and standards may be "important" to most people. The word "annoyed" was used instead to help identify more

			extreme views deemed typical to OC.
LFC	I must always do what I believe is right.	I must do what <i>I</i> believe is right.	Only some participants put emphasis on own beliefs as opposed to majority beliefs when answering the item. “I” was written in bold and italics to emphasize its importance in the phrasing of the item.
LFC	One should always deal with responsibilities before having fun.	One should prioritise responsibilities over having fun.	The word “prioritise” was used by one of the participants and the researcher decided that the phrase better reflects the intended meaning of the question in the context of the theoretical model.
LRO	I purposefully avoid situations in which I could be criticised.	I avoid situations in which I could be criticised.	“Purposefully” was removed to shorten and simplify the question. The word “avoid” already implies doing something on purpose/effortfully.
LRO	I purposefully avoid situations in which I could be seen as weak or incompetent.	I avoid situations in which I could be seen as weak or incompetent.	As above.
LRO	When someone gives me a compliment, I ask myself what it is that they want from me.	When someone gives me a compliment, I ask myself why they are complimenting me.	Participants reported “what it is that they want from me” as too specific but did report wondering why someone compliments them and being suspicious about it at times.
LSC	I rarely feel connected to other people.	I rarely feel deeply connected to other people.	Several participants suggested adding “deeply” or “strongly” to the question to avoid confusion as to what kind of connection the researcher had in mind.
LSC	Being around other people makes me feel alive.	Being around other people makes me feel energised.	Several participants struggled to make meaning of the phrase “makes me feel alive”. One participant suggested replacing it with “energised”.
LSC	When disagreements arise, I withdraw from the situation.	When conflict arises, I withdraw from the situation.	Participants reported that “disagreements” may refer to both debates on certain topics and conflicts/arguments, and that their answers may differ based on the meaning.
LSC	In social situations, I prefer it when people have clearly assigned roles and stick to them.	In social situations, I prefer it when people have clearly assigned roles.	Participants reported that the item is confusing and too long, with most referring mainly to the first part of the question. The question was shortened and thereby simplified.

LSC	I only ever cry on my own or in front of people that I am very close to.	If I were to cry, I would only do it on my own or in front of people that I am very close to.	Several participants reported that they never cry and therefore struggled to relate to the item.
LSC	My relationships with other people are rather superficial.	My relationships with other people are rather shallow.	“Superficial” was replaced with “shallow” due to several participants reporting that they did not understand the word. The choice of the replacement word was consulted with participants who did not have issues understanding the original version of the item.
PIE	I rarely get very excited.	I rarely express my excitement to others.	Several participants suggested that while they may experience excitement on the inside, they rarely express it outwardly, and this is what makes them different from other people.
PIE	I tend to discount and "push down" my emotions.	I tend to dismiss my emotions.	Participants reported that their response may differ if “discount” and “push down” were presented in separate questions. The “push down” part of the question was removed due to another item already dealing with suppression of emotions (“I tend to bottle up my feelings.”). Participants also reported that “dismiss” or “disregard” may be a better choice of word than “discount”. As such, the word was replaced.
PIE	Being emotional is a sign of weakness.	Being emotional makes me feel like I'm a weak person.	Participants reported that they answer to the item would vary depending on whether they were considering others or themselves.
PIE	I am used to being emotionally numb.	Other people feel emotions much more intensely than I do. I don't let my emotions get to me.	Participants reported “numb” to have a strong negative connotation and therefore making them less likely to want to answer to the item truthfully. The also indicated that they may not necessarily feel numb, but just feel emotions less intensely than others, or be less likely to let emotions get to them. As such, the item was replaced with two other items.
PIE	I don't like to reveal my vulnerability.	I don't reveal my vulnerability.	Most participants were referring to not revealing vulnerability rather than not liking to reveal vulnerability. One participant reported becoming confused as while they did not like to reveal their vulnerability, they did so anyway.

PIE	I rarely complain about being stressed or hurt.	I rarely complain about being emotionally hurt. I rarely complain about being distressed.	Participants reported that their answer to the item may differ if “stressed” and “hurt” were presented as separate questions. As such, the two were split into two items, with “hurt” reworded to “emotionally hurt” and “stressed” reworded to “distressed” to better reflect the assumptions of the theoretical model.
PIE	I often feel numb in situations in which others tend to feel intense emotions.	Other people express emotions more openly than I do.	As with the previous item which used the word “numb”, participants reported having negative associations with the word. As such, the item was replaced with a new item, this time aiming to examine open expression of emotions.
PIE	Thinking about crying makes me feel uneasy.	Thinking about myself crying makes me feel uneasy.	Participants reported some confusion as to what the item is trying to get at and suggested that they answers may differ when thinking about themselves crying and thinking about others crying.
PIE	I do not like to dwell on my emotions.	I avoid dwelling on my emotions.	As with one of the previous items, participants reported that while they may not like to dwell, they might dwell anyway, unable to control. As such, “do not like to” was replaced with “avoid”.
PIE	I am a warm and affectionate person.	People see me as a warm and affectionate person.	Participants reported that while they may see themselves as warm and affectionate, other people may not see them in this way.

Next, five items were removed from the item pool ($N_{LFC} = 2$; $N_{LRO} = 1$; $N_{LSC} = 1$; $N_{PIE} = 1$; see Table 5.2).

Table 5.2: Items removed from the item pool based on cognitive interviews with the potential respondents.

Domain	Item	Justification
LFC	I have strong opinions about how things should be done.	Understanding of the item varied between participants, implying ambiguity. E.g., some participants talked about factors that determine whether they should have an opinion, such as financial contributions to the household, and some understood the question in terms of consent. Only some participants understood it as being set in their beliefs about how things should be done. Even then, responses varied between those that reported being set in their beliefs, with participants choosing different response options based on whether they were likely to impose their beliefs on others.

LFC	I must be fair, even if I don't want to be at times.	Participants universally saw fairness as a valued and respected trait, which suggests increased potential for social desirability bias, and thereby low likelihood of accurately discriminating between diverse groups of individuals.
LRO	I strongly believe in the saying "better safe than sorry".	There was considerable variability in participants' interpretation of the phrase "better safe than sorry", with some referring to their personal life and decisions they make (e.g., finances, planning ahead, engaging in risky sports), and some interpreting it as being straightforward with other people so that they do not need to apologise for anything in the future.
LSC	I don't mind starting a conversation with a new person.	Several participants reported the item to be confusing or to be context dependent (i.e., their answer would differ depending on who the new person is and what environment they are meeting the person in). The interpretations of the phrase "new person" varied from a stranger in the street through a new co-worker to a new person at a friendly gathering.
PIE	I rarely laugh out loud.	All participants reported that they do laugh out loud and did not have a lot of thoughts regarding the item, indicating that the item is unlikely to accurately discriminate between groups of individuals.

Additionally, while no direct questions regarding mental health conditions were, a couple of participants who disclosed that they were neurodivergent and provided valuable feedback based on their experience with the questionnaire. They reported that inclusion of absolute words (e.g., "always," "never," "any," "certain," "all") in a questionnaire could be confusing. For example, one participant reported that inclusion of the words "always" in an item has potential to change their answer from *Strongly agree* to *Strongly disagree* if taken literally, as something happening often does not mean that it happens always. As such, absolute words were removed. Additionally, the researcher decided to remove frequency and intensity words (e.g., "often," "strongly," "very") from the items where they were not deemed crucial to the meaning. This was to simplify and shorten the items. Both practices have been previously recommended in best practice guidelines for survey design published by the UK General Medical Council (n.d.). The items to which these changes were made are presented in Table 5.3 below, with any additional changes reported and explained.

Table 5.3: Items from which absolute, frequency, or intensity words have been removed following cognitive interviews with the potential respondents.

Domain	Previous version	New version	Additional changes (if any)
--------	------------------	-------------	-----------------------------

LFC	I like to always be prepared for any possibility.	I like to be prepared for any possibility.	-
LFC	I like to make detailed plans for everything.	I make detailed plans for how to do things.	The word “like” was removed due to the intended focus of the question being on <i>making</i> plans rather than <i>liking to</i> make plans (as per the theoretical model).
LFC	I feel the need to fix all problems immediately.	I feel the need to fix problems immediately.	-
LFC	I strongly dislike it when plans change unexpectedly.	I dislike it when plans change unexpectedly.	-
LFC	I dislike it when my daily routine changes even slightly.	I dislike it when my daily routine is interrupted.	The word “interrupted” was suggested by a participant due to lack of clarity on whether the question pertains to voluntary or involuntary changes to daily routine.
LFC	Structure and order are very important to me.	Structure and order in everyday life are important to me.	While most participants understood the question as intended (i.e., ensuring structure and order in everyday life), some referred to law and legal regulations. To reduce ambiguity, the phrase “in everyday life” was added.
LFC	I pride myself in always appearing to be in complete control.	I pride myself in appearing to be in complete control of my behaviour and emotions.	There was some ambiguity in participants’ understanding of the word “control”, with some referring to coercive control or being possessive of someone else. The phrase “of my behaviour and emotions” was added for clarity.
LRO	I would only ever volunteer to answer a question if I was certain I knew the correct answer.	I would only volunteer to answer a question if I was confident I knew the correct answer.	-
LRO	I am always willing to spontaneously try new things.	I am willing to spontaneously try new things.	-
LRO	I am always on high alert.	I am typically on high alert.	“Typically” was used instead of “always” because it was decided that “I am on high alert.” Would not suffice to reflect the intended meaning of the item.
LRO	I am always happy to get out of my comfort zone and try something new.	I am happy to get out of my comfort zone and try something new.	-

LRO	I would do anything to avoid being the centre of attention.	I avoid being the centre of attention.	-
LRO	Embarrassing myself in front of other people is my worst fear.	Embarrassing myself in front of others is one of my worst fears.	“Other people” was changed to “others” to simplify and shorten the item.
LSC	I am a very sociable person.	I am a sociable person.	-
LSC	I generally prefer being on my own rather than around other people.	I prefer being on my own rather than around other people.	-
LSC	I often yearn to be alone for a while after participating in social events.	I feel the need to be alone for a while after participating in social events.	“Yearn to” was replaced by “feel the need to” due to some participants having difficulty understanding the item/phrase.
LSC	I love meeting new people.	I enjoy meeting new people.	“Love” was pointed out as an absolute and too strong of a word by a participant who disclosed being neurodivergent.
LSC	I tend to hold grudges for a long time.	I hold grudges for a long time.	-
LSC	When other people become upset, I often don’t know how to support them.	When other people become upset, I don’t know how to support them.	-
LSC	I am often misunderstood by other people.	I am misunderstood by other people.	-
PIE	I always choose reason over emotion.	I choose reason over emotion.	-
PIE	I often push through difficult situations without sharing it with anyone.	I push through difficult situations without sharing my struggles with others.	-

The remaining items were retained in the version presented to participants. No further domain changes were made to any of the items due to internal structure analyses being planned at a later stage. The new, 91-item pool ($N_{LFC} = 16$; $N_{LRO} = 23$; $N_{LSC} = 27$; $N_{PIE} = 25$) refined based on the target population cognitive interview study is presented in the Table 5.4 below.

Table 5.4: Refined item pool following target population cognitive interviews.

Domain	Item
LFC	I like to be prepared for any possibility.
LFC	I make detailed plans for how to do things.
LFC	I feel the need to fix problems immediately.

LFC	I dislike it when plans change unexpectedly.
LFC	Structure and order in everyday life are important to me.
LFC	I dislike it when my daily routine is interrupted.
LFC	In day-to-day situations, I rehearse what I want to say over and over again.
LFC	I persist at tasks even if they cause me distress.
LFC	Being correct is more important to me than it is to most people.
LFC	I like my life well-structured and predictable.
LFC	I get annoyed when people don't adhere to social norms and standards.
LFC	Having self-control is more important to me than it is to most people.
LFC	I must do what I believe is right.
LFC	I pride myself in appearing to be in complete control of my behaviour and emotions.
LFC	I cannot stop thinking about a problem until I find a solution.
LFC	One should prioritise responsibilities over having fun.
LRO	I would only volunteer to answer a question if I was confident I knew the correct answer.
LRO	When I notice small discrepancies, I must fix them immediately.
LRO	When someone gives me negative feedback, I just ignore it.
LRO	I carefully consider potential consequences before taking a risk.
LRO	I welcome critical feedback even if I did not ask for it.
LRO	I avoid situations in which I could be criticised.
LRO	I often do things just for fun.
LRO	*I am willing to spontaneously try new things.
LRO	I avoid situations in which I could be seen as weak or incompetent.
LRO	I am typically on high alert.
LRO	*Unpredictability is exciting.
LRO	The world is a scary place.
LRO	When I make a mistake in front of others, I feel more embarrassed than most people would.
LRO	Avoiding negative consequences is a better motivator than gaining something positive.
LRO	I often see a threat where others may not.
LRO	I dislike uncertainty more than most people.
LRO	Life without risks is no fun.
LRO	*I am happy to get out of my comfort zone and try something new.
LRO	I avoid being the centre of attention.
LRO	Embarrassing myself in front of others is one of my worst fears.
LRO	When someone tells me something I do not want to hear about myself, I explain to them why they are wrong about me.
LRO	When someone gives me a compliment, I ask myself why they are complimenting me.
LRO	If people give me positive feedback, it is because they want to manipulate me.
LSC	In social situations, I prefer it when people have clearly assigned roles.
LSC	I rarely feel deeply connected to other people.
LSC	I have few close friends, if any.
LSC	I feel that I am an outsider.
LSC	*Being around other people makes me feel energised.
LSC	*I am a sociable person.
LSC	I prefer being on my own rather than around other people.
LSC	I dislike pointless social interactions.
LSC	I feel the need to be alone for a while after participating in social events.
LSC	When conflict arises, I withdraw from the situation.
LSC	It takes me a long time to warm up to new people.
LSC	I dislike parties.
LSC	*I enjoy meeting new people.
LSC	I often struggle to understand another person's perspective.
LSC	*Being around other people helps me relax.
LSC	I envy people who do better than me.
LSC	If I were to cry, I would only do it on my own or in front of people that I am close to.
LSC	My relationships with other people are rather shallow.
LSC	I behave more awkwardly in social situations than most people do.
LSC	I hold grudges for a long time.
LSC	I struggle to believe in real, unconditional love.
LSC	When other people become upset, I don't know how to support them.

LSC	I dislike being around people who are upset.
LSC	I am different from other people.
LSC	When I see another person crying, I find it difficult to understand why they would do this in public.
LSC	I am distant from other people.
LSC	I am misunderstood by other people.
PIE	I rarely express my excitement to others.
PIE	I tend to dismiss my emotions.
PIE	Being emotional makes me feel like I'm a weak person.
PIE	Other people feel emotions much more intensely than I do.
PIE	I don't let my emotions get to me.
PIE	I choose reason over emotion.
PIE	*I am an open book to other people.
PIE	I push through difficult situations without sharing my struggles with others.
PIE	I don't reveal my vulnerability.
PIE	I rarely complain about being emotionally hurt.
PIE	I rarely complain about being distressed.
PIE	Other people express emotions more openly than I do.
PIE	No matter what I feel on the inside, I make sure I seem fine on the outside.
PIE	Thinking about myself crying makes me feel uneasy.
PIE	My facial expressions don't always match how I really feel (for example, I laugh when I feel awkward, or I smile when I'm sad).
PIE	I consciously put on facial expressions that I think are appropriate in a given situation.
PIE	I tend to bottle up my feelings.
PIE	I tell others that I'm fine, even if I'm not.
PIE	I experience positive emotions strongly.
PIE	I avoid dwelling on my emotions.
PIE	It is difficult for me to stop worrying.
PIE	To support somebody is to help them find a solution to the problem.
PIE	When something bad happens, talking to other people about it helps me a lot.
PIE	People have told me that I'm difficult to read.
PIE	*People see me as a warm and affectionate person.

Note. Reverse-scored items are indicated with an asterisk (*).

The verbal instructions on how to complete the questionnaire given by the researcher at the beginning of each interview appeared to have been well received by participants, with neither individual reporting issues with understanding of how to complete the questionnaire. As such, questionnaire instructions were scripted that included the same information as given to participants during the interview. At times, participants reported feeling as if their answer would depend on the exact context of the situation, they had found themselves in, and as such struggled to pick a definitive response. The researcher paid attention to these items during the analysis to improve clarity. However, the researcher decided to also add a sentence addressing a situation where the respondent may feel that no answer exactly reflects who they are. The final instructions are presented below:

“Please review the instructions carefully before filling out the questionnaire. Below you will see a list of questions about how you typically behave and respond to situations in everyday life. There are

seven possible answers for each question: (1) Strongly disagree, (2) Disagree, (3) Slightly disagree, (4) Slightly agree, (5) Agree, (6) Strongly agree, and (7) No opinion. **Please read each question and select the answer that best describes you as a person.** While you may not find a response that exactly reflects who you are, try to pick the one that is the closest to how you would describe yourself. There are no right or wrong answers – try to go with the answer that first comes to mind.”

5.4 Discussion and conclusions

The initial aims of the study were to assess how well the item pool reflects the construct of interest (content validation) and ensure coherent understanding of the items between participants and semantic simplicity and clarity of the items in the pool (pre-testing). For this purpose, a two-stage study was designed. In the first stage, a sample of participants was screened for overcontrolled tendencies using quantitative methods. In the second stage, cognitive interviews were conducted with individuals whose screening scores indicated high overcontrolled tendencies.

In Stage 1 of the study, the ASC-WP questionnaire designed in accordance with Lynch’s (2018) theoretical model that guided item development for the new measure of overcontrol was used for screening of participants. The same questionnaire was used for screening of participants in the inductive item development study (see Chapter 3). Findings of the item development study suggested that the questionnaire may overestimate overcontrolled tendencies of respondents. Because one of the initial aims of the current study was content validation via target population judgement, it was important that cognitive interview participants were indeed highly overcontrolled. To increase the likelihood that the high scores on the ASC-WP indicating overcontrolled tendencies were accurate, the researcher decided to exclusively recruit participants from mental health groups on social media. The rationale behind this was that people who participate in such groups are more likely to have had ill mental health linked to under- or overcontrolled personality.

Once cognitive interviews began, the researcher observed that for several participants response patterns were not congruent with theoretical descriptions of highly overcontrolled individuals. This again suggested the ASC-WP may be

overidentifying overcontrol, and that purposive recruitment from social media groups and forums concerning mental health was not sufficient in preventing potential misidentification of overcontrolled tendencies. As a result, the researcher was not able to confidently claim that the screening measure accurately identified participants with overcontrolled tendencies, despite targeted recruitment method introduced. Nevertheless, the data produced during interviews provided useful insights into whether participants understood the items in the pool as intended by the researcher, and whether they were written in a clear, straightforward way. As such, a decision was made to abandon the goal of further validating the contents of the scale, and only use the cognitive interviews data only for the aim of pre-testing. This is because a relatively high degree of certainty that the participants are characterised by overcontrol would be required to be able to use the collected data for the purpose of confirming how the construct manifests in the population of interest. Forsaking the content validation aim of the study was not a major concern, as content validity study was already conducted with expert judges with satisfactory results (see Chapter 4).

Further, for some participants, there were also discrepancies between demographics self-reported during the screening stage and demographics observed during the interview. This could indicate that participants did not pay sufficient attention when responding to questions during the screening stage, potentially not only in the demographic questionnaire, but also the items of the ASC-WP. This could provide an alternate explanation of why participants with high scores on the ASC-WP questionnaire did not respond to items of the newly developed questionnaire in a manner that would be expected of a highly overcontrolled individual. Alternatively, there was a possibility that some participants relied on AI tools in completing the questionnaire.

Another issue that occurred was that a less-than-optimal number of participants was available for the cognitive interviews due to several participants not booking an interview following an invitation and the number of no-shows among those who did book an interview. Due to this, as well as the other issues experienced during the screening stage in the first round of recruitment that were previously discussed, a decision was made to conduct a second round of recruitment. This time, the researcher decided to opt for a convenience sample and approach individuals from their social

network. Convenience sampling in research is both widely used due to its opportuneness and ease of recruitment and widely criticised due to the bias of the sample reducing generalizability of the results (Emerson, 2021). In the current study, only part of the sample was recruited via convenience sampling, hopefully offsetting the potential sample bias issues. The same cut-off score for overcontrolled tendencies was also used in the screening stage to ensure consistency between the samples recruited in both rounds. Further, where participants seemed to have similar doubts about the items as the participants before them, they were asked additional questions that helped verify whether their conclusions are indeed the same or very similar to those of previous participants. For instance, where a participant hesitated or changed their answer to a particular item, but was struggling to voice why, the researcher would ask a direct question based on the previous participants' answers. For example, for the item "I rarely complain about being stressed or hurt," an unsure participant was asked whether their hesitance is due to the words 'stressed' and 'hurt' being interpreted very different. This was based on previous, congruent responses of several respondents interviewed beforehand, and indeed seemed to help the participant arrive at their answer, simultaneously further verifying previous responses. The researcher ensured to ask the iterative questions in a way that was not overly suggestive, allowing the participant to still be able to comfortably disagree with the interviewer.

With online surveys were growing in popularity, several challenges of quantitative research conducted online encountered in the current study have been previously discussed in literature due to concerns about potential low-quality data (Shamon & Berning, 2019). The risk for low quality online survey responses is especially high in cases where incentives are offered to participants (Shamon & Berning, 2019). This is because, motivated by compensation, participants may be more likely to, for instance, participate in the study multiple times (Teitcher et al., 2015), respond carelessly to get to the end of the survey as quickly as possible (Shamon & Berning, 2019) or, more recently, use AI assistance (i.e., bots) to complete surveys for them (Xu et al., 2022). Because of the current study using an online survey only as a screening measure and offering incentives only following completion of the interview stage, the researcher did not anticipate bot or repeated responses to be an issue, especially following a largely successful item generation study conducted in a similar way. Nevertheless, the inconsistencies in the demographic data and a high drop-out

rate may suggest that some study participants were largely motivated by incentives, paid little attention to the items, tried to participate in the study repeatedly, or used survey bots to complete the screening questionnaire.

While the described issues may indicate reduced reliability of the demographic data collected during the screening stage (which should be interpreted with caution), due to the study design utilising mixed methods and the aim of further content validation abandoned, the researcher was still able to ensure that data collected during the cognitive interviews was viable to use for the purpose of pre-testing the newly developed scale. This is because the interviews were video- and audio-recorded, allowing the researcher to ensure participants' engagement in the study, as well as that no individual participated in the study more than once. Having a second, known sample of participants increased the likelihood of problem detection and allowed the scale developer to, through iterative questioning, ensure that the responses indicating issues were congruent across the two samples, adding to the reliability of the collected qualitative data (Blair & Conrad, 2011). Reflecting upon the sampling issues encountered in the current study, a decision was made to, going forward, include anti-bot captcha and attention-check questions in the quantitative elements of studies as a precautionary measure.

6. Internal structure and reliability of the OAQ

6.1 Introduction

Once the content of the item pool has been content-validated and the items pre-tested, the next steps in the scale development process are dimensionality analyses, factor retention, and item reduction. According to Boateng and colleagues (2018), dimensionality analyses are used to extract the internal structure of the scale. Factor retention analyses then determine the number of factors that is the most optimal for a given set of items and understand the relationships between the items that comprise the factors based on shared variance. The goal of the item reduction, in turn, is to ensure that the final item pool only includes functional, internally consistent items that accurately represent the domain under study. The structure of the scale should also be verified either with the same sample of participants at a different time point, or, ideally, with a different sample of participants.

Methods that may be used to establish the optimal factor structure and number of items are factor analyses. Factor analyses are multivariate statistical techniques commonly employed for variable reduction and dimension identification (Hair et al., 2019; Rennie, 1997). Importantly, they allow the researcher to develop⁸ and evaluate the construct validity of a new scale – i.e., ensure that the respondents' scores on the scale offer meaningful information about the construct that the scale developer intended to measure, in the way that it was operationalised (Colton & Covert, 2007; Murphy & Davidshofer, 2005). Namely, factor analyses allow to assess structural accuracy of the measure – the extent to which the relationships between the items in the pool accurately represent the conceptual framework (Hughes, 2018). Further, at this stage, preliminary reliability evidence is also obtained, seeking to establish the internal consistency of the scale – the extent to which items designed to assess a single construct correlate with each other to consistently measure that construct under similar conditions in a particular population (Dunn et al., 2014; Hair et al., 2019; Murphy & Davidshofer, 2005).

⁸ For a discussion on why construct validity is *developed*, and not *established* or *determined*, see Strauss & Smith (2009).

The paragraphs that follow discuss the two factor-analytic statistical techniques used in scale development – the Exploratory Factor Analysis (EFA) and the Confirmatory Factor Analysis (CFA) – and compare their aims, methods, and utility in developing new psychometric measures. Methodological issues relevant to factor-analytic studies are also discussed, including sample size requirements, model estimation methods, factor retention methods, and methods for obtaining preliminary reliability evidence concerning the internal consistency of the newly developed scale.

6.1.1 Exploratory factor analysis (EFA).

EFA is one of several exploratory techniques for multivariate analysis that can be used in scale development, the aim of which is two-fold (Hair et al., 2019):

- (1) to condense the variables – that is, in measure development, the scale items – into factors to define the structure of the latent construct (data summarisation),
- (2) to reduce the number of items to a most parsimonious set possible in a way that allows for a minimal loss of information (data reduction).

EFA groups together items that assess the most similar content and are the most highly interrelated into factors – dimensions that underlie the data (Hair et al., 2019). Importantly, EFA is data-driven – it derives the factors that best reflect the latent construct directly from the data and does not require the scale developer to have any preconceptions about the possible structure of the scale (Murphy & Davidshofer, 2005). The technique allows for the items to freely load on any factor, without the need to indicate which item was intended to belong to which factor (Worthington & Whittaker, 2006).

In case of Likert-type scales with more than two response options, the relationships between items in EFA are typically determined based on polychoric (inter-item) correlation matrices. According to Fabrigar and colleagues (1999), the correlation matrices help to understand the patterns of correlations between the items and the items and common factors. The latter is indexed by factor loadings (extracted from the correlation matrices), which indicate how strongly the items load on the underlying factors. Factor loadings below .30 or .40 are typically considered suboptimal and indicate the corresponding items as candidates for removal, with larger loadings necessary when sample sizes are relatively small and the number of factors in

the model is greater (Boateng et al., 2018; Hair et al., 2019; Williams et al., 2010). Moreover, items that have a high loading on more than one factor (i.e., cross-loadings $> .30$ or $.40$) should also be considered for deletion so that a model structure as simple as possible can be achieved (Hair et al., 2019). As factor loadings indicate how well the items represent the underlying construct, they also provide information about construct validity of the scale.

Notably, EFA requires the scale developer to specify the intended number of factors to be included within the model in advance (Gerbing & Hamilton, 1996). According to Hayton and colleagues (2004), how many factors to include is a decision critically important in conducting the analysis. This is because both under- and overestimation of the number of factors can be detrimental to the reliability of the results. Overestimation of the number of factors introduces the risk of overemphasising the importance of minor factors that may be difficult to replicate in future studies. Underestimation, in turn, leads to loss of meaningful information that would allow to better measure and understand the latent construct due to distorted factor loadings.

Several techniques exist that help scale developed make decisions on factor retention. The most known is the Kaiser's (1970) rule (of thumb), which indicates that all factors with eigenvalues > 1 should be retained within the model. However, as Hayton and colleagues (2004) explained in their paper, the method is population-based rather than sample-based (i.e., its assumptions are based on a population of an infinite size; Dinno, 2009) – and therefore prone to sampling errors that often result in overestimating the number of factors due to inflated eigenvalues. One popular alternative to the Kaiser's rule is parallel analysis (Horn, 1965) – a method that considers both the sample size and the number of variables, adjusts for the sampling error by comparing observed eigenvalues to simulated eigenvalues, and establishes a factor retention threshold (Hair et al., 2019; Thompson & Daniel, 1996; Turner, 1998). Based on evaluation studies, parallel analysis as a factor retention method yields impressive results compared to other available methods (e.g., Velicer et al., 2000; Silverstein, 1987; Zwick & Velicer, 1986) and has been described by Hayton and colleagues as “one of the most accurate factor retention methods” (p. 191).

6.1.2 Confirmatory factor analysis (CFA).

CFA is another multivariate statistical method used in scale development. It is a type of structural equation modelling concerned with the relationships between observed variables (i.e., test items and tests scores) and latent constructs or factors (Brown, 2015). As the name indicates, it is confirmatory rather than exploratory in nature, and theory-driven – it is used either as an alternative to an exploratory technique when there is pre-existing evidence for a specific model structure, or a supplement to another multivariate analysis when the scale developer wishes to extend the validity evidence for a previously established model (Gerbing & Hamilton, 1996; Hair et al., 2019; Harrington, 2009). Alike EFA, CFA also returns factor loadings, based on which construct validity evidence can be obtained. Moreover, due to its theory-driven nature, good CFA model fit also supports the construct validity of the scale.

An important function of CFA in scale development is to assess (and help improve) the goodness of fit of the theorised model specified by the scale developer. The values of both absolute fit indices (which indicate how well the theorised model fits the collected data) and incremental fit indices (which indicate the fit of the theorised model compared to a baseline model) are used as indicators of good or poor model fit (Hair et al., 2019). The most widely used absolute fit indices include the Chi-squared (χ^2) statistic, the Root Mean Square Error of Approximation (RMSEA), and the Standardised Root Mean Square Residual (SRMR). The most reported incremental fit indices include the Tucker Lewis Index (TLI) and the Comparative Fit Index (CFI). According to an overview of fit indices by Hair and colleagues (2019):

- The χ^2 fit index is a test of statistical significance, with lower χ^2 values and non-significant p -values indicating higher concordance between the theorised model and the observed model – i.e., a good model fit. Notably, the χ^2 index is sensitive to the number of variables and the sample size, with greater numbers of variables and a smaller number of respondents making a significant result more likely. As such, χ^2 is rarely used as to determine fit when conducting CFA.
- The RMSEA aims to address the limitations of the χ^2 test by adjusting for a large number of variables is large or a small sample size. It is more population-based than sample-based. The values of the RMSEA range from 0 to 1, with lower values indicating better model fit.

- The SRMR compares the predicted and observed models based on the standardised value of an average of the residuals (the standardised average of errors when predicting covariance terms). It is sensitive to the number of variables and respondents, Lower values of the SRMR indicate better model fit, with values $> .10$ typically considered to indicate poor fit.
- The TLI compares χ^2 values of the theorised model and a baseline model. It is somewhat sensitive to model complexity. Its values can fall below 0 and rise above 1, as it is not a normed fit index. The TLI values approaching 1 indicate a good fit.
- The CFI is the most widely used incremental fit index that is normed and insensitive to model complexity. It also compares χ^2 values of the theorised model and a baseline model. Its values range from 0 to 1, with values closer to 1 indicating better model fit.

Over the years, there has been much discussion on recommended cut-off values of fit indices other than the χ^2 for the model to be considered a good fit. The most used cut-off values are ones published by Hu and Bentler (1999) – with TLI and CFI $> .90$ and the RMSEA and SRMR $< .08$ indicating an acceptable fit, and TLI and CFI $> .95$ and the RMSEA and SRMR $< .06$ indicating a good fit. Notably, however, these cut-off values were based on a ML estimation method and a relatively simple model. Nevertheless, in the absence of a simple method of obtaining cut-offs for fit indices, these values are also commonly applied in scenarios where ordinal data with other model estimators are used, non-normal data is present, and more complex models are specified (see e.g., Finney & DiStefano, 2013; Padgett & Morgan, 2020). It is therefore commonly stated that there are no gold standard cut-off values when evaluating goodness of fit (see e.g., Hair et al., 2019; Marsh et al., 2009). In search for an alternative solution, a dynamic fit indices website is currently being developed (McNeish & Wolf, 2023, 2024) that would allow for a more accurate estimation of appropriate fit indices for specific models based on Monte Carlo simulations. However, it is currently only available for a limited number of model structures. With no simple solution yet available, Hair and colleagues (2019) provided *flexible* guidelines for use of fit indices that consider the sample size and the number of variables in the model. For example, for sample sizes > 250 and between 12 and 30 variables, they indicated that for a good-

fitting model, χ^2 *p*-values may still be significant, but with CFI and TLI > .94 and the SRMR < .08 and the RMSEA < .07. They also noted, however, that “(t)he desire to achieve good fit should never compromise the theory being tested. (...) Researchers learn not only from theory that is confirmed, but also from the areas where theoretical expectations are not confirmed” (p. 651). They recommended against reducing the number of items per (sub)scale to two or three items in pursuit of achieving a better model fit, in order not to compromise the theoretical validity of the measurement instrument and suggested that the fit of different theoretical models should be compared where possible.

6.1.4 Methodological considerations for conducting factor analyses.

Model estimation and factor rotation methods.

When conducting factor analyses, the scale developer must select a suitable model estimation method for the type of data analysed to achieve the most accurate results regarding the factor structure, to reliably examine the validity of the construct being measured (Kyriazos & Poga-Kyriazou, 2023). The maximum likelihood (ML) estimator is the most widely used estimator, despite being intended for use with continuous rather than ordinal data and assumes multivariate normality – a rarity in psychological research (Kořar & Kořar, 2015; Kyriazos & Poga-Kyriazou, 2023). As such, use of estimation methods that are appropriate for use with ordinal data is recommended – i.e., methods that use a least squares approach to factor analysis – that are less sensitive to non-normality (e.g., the minimum residuals in EFA and weighted least squares with means and variance adjusted in CFA; Asparouhov & Muthén, 2010; see e.g., Cornejo et al., 2021).

Further, factor analyses require the scale developer to specify an appropriate rotation method for models with more than one factor (Yang & Xia, 2015). The aim of rotation is to obtain meaningful factors with the simplest possible factor structure (Hair et al., 2019). There are two types of rotation – orthogonal and oblique. Orthogonal rotation assumes that the factors are independent from each other, while oblique rotation allows for factor correlations (Rennie, 1997). Orthogonal rotation methods (e.g., varimax, quartimax, equimax) are considered most appropriate when the factors are not expected to be correlated. However, when the main goal of the analysis is to identify factors that are theoretically meaningful and expected to be correlated with

each other, oblique rotation methods (e.g., oblimin, promax) are more suitable (Hair et al., 2019; Rennie, 1997).

Sample size in factor-analytic studies.

An important consideration when using factor analyses, whether exploratory or confirmatory, is also the sample size. Analyses that include many variables require large statistical power, and as such, large sample sizes. While recommendations as to sample size for factor analyses vary widely in the literature, a commonly reported rule of thumb is that the ratio of participants to variables should be at least 5:1 – albeit with some arguing that a ratio of 10:1 or even 20:1 is much more likely to provide reliable results (Hair et al., 2019). Moreover, when using both exploratory and confirmatory multivariate techniques, a separate sample of this size is strongly recommended for each of the analyses to avoid an overestimation of the model fit and estimation parameters (see e.g., Fokkema & Greiff, 2017).

Choosing the right approach.

Factor analyses can be used both in scale development and scale evaluation studies (Rennie, 1997). The choice between exploratory and confirmatory approaches to factor analysis in scale development typically depends on the subject being studied – i.e., traditionally, it depends on how much is known about the latent construct that the scale is intended to measure (Murphy & Davidshofer, 2005). Due to its exploratory nature, EFA is used primarily in cases where there are no a priori assumptions about the structure of the scale, based on, for example, existing theory or research. In cases where there is an existing hypothesis about the structure of the data and the goal is to test this hypothesis, confirmatory approaches – such as CFA – are typically used instead (Hair et al., 2019; Hurley et al., 1997; Murphy & Davidshofer, 2005).

However, it is often highlighted that confirmatory approaches are the right choice only when there is a *strong* foundation for the hypothesised structure – and where there is little evidence, more preliminary work ought to be conducted first (Harrington, 2009). While confirmatory approaches are considered as generally more powerful, exploratory approaches are argued to be the appropriate choice when there is not enough information about a construct to confidently form hypotheses (Hurley et al., 1997; Murphy & Davidshofer, 2005). Moreover, some researchers also

recommended for exploratory techniques to be used as a precursor to confirmatory techniques as means of cross-validating the results (Gerbing & Hamilton, 1996). Hurley and colleagues (1997) even argued that in initial scale development, using CFA may be counterproductive, because it is unlikely to fit the data well so early in the process – sometimes even after the most appropriate items have first been selected using EFA.

6.1.5 Assessing internal consistency of psychometric scales.

Factor analyses allow for extraction of preliminary evidence regarding the reliability of the newly developed scale and its subscales (Brown, 2015). In terms of psychometrics scales, there are two types of reliability that are typically considered: (1) the consistency of test scores (e.g., test-retest, alternate forms, or split-half reliability), and (2) consistency of test contents (internal consistency; Murphy & Davidshofer, 2005). Reliability in terms of the consistency of test scores over time (specifically, the test-retest reliability) is discussed in more detail in the next chapter. In the current chapter, the latter type of reliability – known as internal consistency – is considered.

As previously stated, internal consistency is concerned with the extent to which multi-item (sub)scales, designed to assess a single construct correlate with each other, consistently measure that construct under similar conditions in a particular population (Dunn et al., 2014; Hair et al., 2019; Murphy & Davidshofer, 2005). The most widely used index of internal consistency is Cronbach's (1951) coefficient alpha (McNeish, 2018; Padilla, 2019; Peterson and Kim, 2013). It is computed based on the values of item inter-correlations in the covariance matrix and the number of items (Murphy & Davidshofer, 2005). Cronbach's alpha values $> .70$ are widely considered to indicate satisfactory internal consistency, and very high values have been suggested to indicate potential item redundancy (Taber, 2018; Tavakol & Dennick, 2011).

However, despite Cronbach's alpha's widespread use, there are drawbacks to using the coefficient to assess the internal consistency of Likert-type scales (see e.g., Cortina, 1993; Dunn et al., 2014; Peterson, 1994; Schmitt, 1996). Cronbach's alpha makes several assumptions that must be met for the coefficient to produce reliable values, that are often overlooked by its users (Kalkbrenner, 2023). For instance,

Cronbach's alpha assumes that the responses to items are continuous, normally distributed, and tau-equivalent (i.e., that the true score variance is constant, which is rarely the case in psychological scales; Dunn et al., 2014; Zumbo et al., 2007). This can often lead to an underestimation of the true reliability, especially for tests where there are less than five response options (Osburn, 2000; Peterson & Kim, 2013; Zumbo et al., 2007).

Aiming to offset some of these limitations, Zumbo and colleagues (2007) introduced an alternative, less conservative coefficient alpha designed to examine ordinal (sub)scales – known as the ordinal coefficient alpha. Ordinal coefficient alpha relies on a polychoric correlation matrix when estimating internal consistency, rather than a covariance matrix. In Zumbo and colleagues' simulation study, the coefficient has been shown to be a dependable predictor of internal consistency of ordinal data – regardless of the number of response options, skewness of the data distribution, or the degree of theoretical reliability. These results suggest that the ordinal coefficient alpha is a more viable choice than Cronbach's alpha when examining internal consistency of Likert-type scales. However, like Cronbach's coefficient alpha, the ordinal coefficient alpha does assume tau-equivalence, which could introduce issues in reliably assessing the internal consistency of scales. The performance of ordinal alpha when assumptions are violated is yet to be extensively explored (Kalkbrenner, 2023).

A distinct type of internal consistency that does not assume tau-equivalence is composite reliability. It is used in structural equation modelling studies, such as those using CFAs, and uses factor loadings and uniqueness to estimate true reliability of scales (Dunn et al., 2014; Hair et al., 2019; Padilla & Divers, 2016; Peterson & Kim, 2013). One of the most prominent estimates of composite reliability that can be reliably used with ordinal data is McDonald's (1999) omega (Kalkbrenner, 2023). Dunn and colleagues (2014) reported that omega makes for a more sensible choice when assessing the internal consistency of psychological scales, as it makes fewer and more realistic assumptions, and is less likely to cause inflation- and attenuation-related issues. However, analyses using structural equation models can be quite complex and typically require relatively large sample sizes, which can be an issue for resource-limited projects (Viladrich et al., 2017; Yang & Green, 2010). As to acceptable thresholds for omega, the numbers are roughly the same as for alpha. Hair and

colleagues (2019) indicated that composite reliability between .70 and .95 (or between .60 and .95 in exploratory research) indicates satisfactory to good internal consistency – with scores > .95 indicating that item redundancy is likely.

Both ordinal alpha and omega, however, assume the data to be unidimensional (Kalkbrenner, 2023) – and therefore their use is most appropriate for assessing the composite reliability of unidimensional scales or unidimensional subscales of multidimensional scales. To assess composite reliability of multidimensional scales, different types of coefficient omega have been proposed. A hierarchical omega coefficient estimates the composite reliability of a general factor comprised two or more smaller, more specific factors – but not necessarily adhering to a higher-order structure (Flora, 2020). A type of omega that is used specifically in assessing composite reliability of scales with a higher-order structure is higher-order omega coefficient which estimates the composite reliability of the higher-order factor (Flora, 2020). Rather than alpha or standard omega coefficients, hierarchical or higher-order omega (depending on the factor structure) should be used to assess the composite reliability of general factors in multidimensional scales when the resources allow for structural equation modelling analyses.

6.1.6 The current studies.

This chapter will present two factor-analytic studies – one using an exploratory, and one using a confirmatory approach. The aim of the studies is to develop an internally reliable and construct-valid scale that assesses the four deficits of maladaptive overcontrol as conceptualised by Lynch (2018).

The researcher decided to precede CFA with EFA despite existing preconceptions about the possible structure. It is the researcher's view that EFA can be useful when some *a priori* assumptions have been made, but the theory based on which they were developed is, for example, relatively novel, and does not yet have a strong empirical grounding – as is the case in the current thesis. As such, a decision was made, in the current thesis, to utilise both EFA and CFA. EFA will be used for the purpose of selecting only the most appropriate items and factors that will explain maladaptive overcontrol in the context of the four core deficits (Lynch, 2018). CFA will be used to

cross-validate the results and establish the goodness-of-fit with Lynch's theorising on the four core deficits on maladaptive overcontrol.

Considering the sample size recommendations, the number of items in the version of the measure to be used ($N = 91$), and the resources available to the researcher, 600 responses will be collected for each the EFA and the CFA to exceed the recommended minimum participant-to-item ratio of 5:1. Due to time restrictions, data collection for both studies will take place at the same time, using the 91-item version of the OAQ. For the CFA, the items rejected based on the EFA will be deleted from the CFA dataset before the analysis will be conducted. Factor rotation methods and model estimators appropriate for use with ordinal data will be used throughout. Flexible guidelines for interpreting CFA goodness-of-fit indices provided by Hair and colleagues (2019) will be applied, and a conscious effort will be made not to compromise the theoretical validity of the OAQ in pursuit of better model fit. However, due to lack of empirical research on maladaptive overcontrol as defined by Lynch (2018) and, thereby, alternative theoretical models, no model comparisons will be performed. Once a final, higher-order CFA model of the OAQ is established to evaluate the internal consistency (including composite reliability), both ordinal alpha and omega values will be calculated at subscale level, and the higher-order omega coefficient will be calculated at a global level.

6.2 Method

The study plan was submitted for review by the Schools of Business, Law and Social Sciences Research Ethics Committee (reference ID: 1536443). The application included documents such as participant information sheet, consent forms, and debrief forms – all of which were developed in accordance with the British Psychological Society guidelines and internal Nottingham Trent University guidelines. A favourable ethical opinion was granted for the conduct of the study.

6.2.1 Design.

The overarching aim of the two studies was to develop and examine the construct validity of the OAQ. To achieve that, two factor-analytic studies were conducted, in which the internal psychometric properties of the new scale were evaluated and improved:

Study 1. EFA was conducted to: (1) ensure that items that were irrelevant to the conceptual framework were identified and removed from the item pool; and (2) examine and refine the internal structure of the scale and its subscales to ensure their consistency with the theoretical perspective driving the scale development.

Study 2. CFA was conducted to cross-validate the internal structure of the OAQ, examine the fit of the hypothesised model of maladaptive overcontrol to the data, and further refine the scale. The internal consistency of the OAQ was also assessed.

6.2.2 Participants.

To take part, the participants needed to be 18 years old or over and fluent in English, and resided in English-speaking countries that fall within the Western, Educated, Industrialised, Rich, and Democratic (WEIRD) framework – the United Kingdom (UK), the United States of America (USA), Ireland, Canada, Australia, and New Zealand. While for the moment literature does not indicate sex differences in overcontrol, the researcher believed that for the results to be widely generalisable, the data should be as balanced as possible. As such, to ensure even split of participant sex (at birth), the balanced sample option was chosen on Prolific, allowing for collection of 50% female and 50% male responses.

After collecting 615 valid responses out of the intended 1200, the study was paused, and the collected data was inspected by the researcher. Data inspection revealed that a vast majority of participants ($N = 527$, 85.69%) were UK residents. To account for UK oversampling, decision was made to pause recruitment and run a separate survey that would recruit only from the USA, Canada, Ireland, Australia, and New Zealand to account for UK oversampling. However, this meant that the sample from the first round of recruitment was not balanced in sex, with 431 of the 615 collected valid responses (70%) were completed by females. Still, to recruit as many male participants as possible, the second round of recruitment also implemented the balanced sample option on Prolific.

For a response to be valid and approved, the participants must have consented to take part in the study, indicated that they were residing one of the countries of interest, that they spoke fluent English, successfully submitted the survey, and failed

no more than one attention check (three attention checks were included; e.g., *This is an attention check, please select 'No opinion' as an answer to this question*). In both rounds of recruitment, a total of 1200 valid responses were collected (615 in the first round, and 585 in the second round). Each participant who submitted a valid response was rewarded £1.50 for their time⁹.

The sample included 723 females (60.25%) and 469 males (39.08%); five participants reported their sex at birth as Other (.25%), and five people chose not to disclose their sex (.42%). The average age of participants was 39.04 years old ($SD = 13.52$, range: 18-86 years old). Five hundred and twenty-seven participants resided in the UK (43.92%), 229 in Canada (19.08%), 216 in the USA (18.00%), 124 in Australia (10.33%), 56 in Ireland (4.67%), and 48 in New Zealand (4.00%). All participants declared fluency in English, with it being the first language for 1079 participants (89.92%) and a preferred language for 110 participants (9.17%). For more information on the demographic characteristics of the sample, please see Appendix G. The total sample of 1,200 was split in half, with the intention to use one part in the EFA study and the other part in the CFA study. For details on sample split and characteristics of the samples, see section 6.3.1 of the current chapter.

6.2.3 Materials.

The survey was administered to participants using Qualtrics and Prolific platforms. It included the newly developed measure of overcontrol, i.e., the 91-item version refined based on the cognitive interviews. A demographic questionnaire was also included, with questions regarding the following variables: age, country of origin, country of residence, ethnicity, sex assigned at birth, gender identity, marital status, and education level. It also included a question as to whether English was a participant's first or preferred language.

6.2.4 Procedure.

The survey was advertised exclusively on Prolific and filled in by participants using Qualtrics. Participants were first presented with a *captcha* to screen for AI-assisted entries. Upon successful completion of the scan, participants were presented with an

⁹ This is in line with the Prolific policy regarding a minimum payment of £6.00ph pro rata. Approximate length of the study was estimated based on Qualtrics predictions.

information sheet, a consent form, and a demographic screener. If a participant did not consent to taking part in the study, indicated residing in a country other than the UK, USA, Ireland, Canada, Australia, or New Zealand, and/or indicated that they did not speak English fluently, they were automatically redirected back to Prolific. On Qualtrics, after consenting to take part and filling in the demographic information, participants were presented with the 91-item version of the newly developed measure of overcontrol, including instructions on how to answer the items. Questionnaire items were randomised across domains (i.e., questions from all domains were intermixed) and participants (i.e., the order of questions differed between individuals). This was to mitigate order effects. Three attention checks were included within the questionnaire. As per Prolific policy, two or more failed attention checks resulted in rejection of the submission. After completing the questionnaire, participants were presented with a completion code necessary to confirm study participation and a link to the debrief file download. All participants who submitted the completed survey on Qualtrics, confirmed their participation on Prolific, and did not fail more than one attention check were approved, and their responses were treated as valid for analysis. Data cleaning, data split, and data analysis procedures are described in detail in the sections to follow.

6.3 Data collation

Microsoft Excel and RStudio were used for data collation and cleaning. RStudio was used for data split. All other statistical analyses were also conducted using RStudio. RStudio software in version 2023.06.0+421 was running under macOS Ventura 13.0.1. Packages used included base (v4.2.2), stats (v4.2.2), naniar (v1.1.0), tidyverse (v2.0.0), psyntur (v0.1.0), nFactors (v2.4.1.1), mice (v3.16.0), psych (v2.2.9), lavaan (v0.6-16), and semTools (v0.5-6).

6.3.1 Data cleaning.

Demographic data.

Participants' answers to all demographic questions were reviewed. To look for any additional categories that might have come up repeatedly, the researcher inspected all responses where participants chose "Other" as a response option and inserted their own descriptor instead. Several participants ($N = 12$, 1%) who chose "Other" as a response when asked about their sexual orientation indicated being asexual. As such, a separate

“Asexual” category was added for the *Sexual orientation* variable. For the text-entry *Country of origin* variable, to ensure consistency among country names, all values entered by participant were reviewed and any variations, aliases, and spelling errors were addressed (e.g., “United Kingdom,” “U.K.,” “Uk,” and any other variations on the name were recoded to “UK”).

Psychometric data.

Textual labels for numerical response options to the newly developed questionnaire were recoded as numerical values (*Strongly disagree* – 1, *Disagree* – 2, *Slightly disagree* – 3, *Slightly agree* – 4, *Agree* – 5, *Strongly agree* – 6). All *No opinion* responses ($N = 3024$, 2.85%) were recoded as missing values in accordance with the intended purpose of the response option. Of note, there were 16 missing responses (0.0002%) to the questionnaire where no response option has been chosen (missing values $N_{\text{total}} = 3040$, 2.86%). Numerical response options to all reverse-scored items across the four domains ($N_{\text{total}} = 14$, 15.38%; $N_{\text{LFC}} = 0$, 0.00%; $N_{\text{LRO}} = 6$, 26.09%; $N_{\text{LSC}} = 4$, 14.81%; $N_{\text{PIE}} = 4$, 16.00%) were recoded from 1:6 to 6:1 to standardise the response scales.

6.3.2 Missing data

There were 3040 missing datapoints in the psychometric data, accounting for 2.78% of all values. Little’s Missing Completely at Random (MCAR) test (Little, 1988) was conducted to inspect the missing data patterns. The results indicated no significant differences between the patterns ($\chi^2(35709) = 984.81$, $p > .999$), and thereby provided no evidence against the hypothesis that the data are MCAR.

6.3.3 Data split.

The total sample was split to create two separate samples of 600 participants – one sample for use in Study 1 (EFA), and one sample to use in Study 2 (CFA). Using a random split function in RStudio, the rows in dataset containing data collected from the total sample of participants were randomly shuffled, and the sample was split into Study 1 dataset (rows 1:600) and Study 2 dataset (rows 601:1200).

6.4 Demographic characteristics: Data analysis and results

The differences between demographic characteristics of Study 1 and Study 2 samples were compared. An independent sample t-test was used to compare the samples in age with no significant difference between the groups (see Table 6.1). Chi-square tests with Monte Carlo simulation¹⁰ based on 2000 replicates were conducted to compare the differences in categorical demographic characteristics, and no significant differences were found (see Table 6.2 and Appendix H).

Table 6.1: Differences in age between EFA and CFA participant samples following sample split.

Variable		EFA		CFA	<i>t</i>	<i>df</i>	<i>p</i>
Age (in years)	<i>M</i>	39.34	<i>M</i>	38.76	.75	1194.40	.455
	<i>SD</i>	13.64	<i>SD</i>	13.41			
	<i>Min.</i>	18	<i>Min.</i>	18			
	<i>Max.</i>	86	<i>Max.</i>	76			

Table 6.2: Differences in selected demographic characteristics (categorical) between EFA and CFA participant samples following sample split.

Variable	Category	<i>N (EFA)</i>	<i>N (CFA)</i>	χ^2	<i>p</i>
Residence	Australia	65	59	28.73	.268
	Canada	108	121		
	Ireland	25	31		
	New Zealand	22	26		
	UK	268	259		
	USA	112	104		
English language	First	540	539	8.58	.179
	Preferred	52	58		
	Neither	8	2		
	Not reported	-	1		
Sex	Female	367	356	2.57	.863
	Male	230	239		
	Other	1	2		
	Not reported	2	3		

Note. Degrees of freedom were not determined due to *p*-values being obtained through Monte Carlo simulation.

¹⁰ Monte Carlo simulation was utilised to account for some cell counts <5, with Chi-square tests with opted for over Fisher's exact tests due to large sample size.

6.5 Exploratory Factor Analysis (EFA): Data analysis and results

The *minres* estimation method was used when conducting EFA analyses throughout this study. Where more than one dimension was indicated, *promax* rotation method was used. This was because oblique rotations allow for factor correlations (Costello & Osborne, 2005), which were expected based on the theoretical model. The pairwise deletion was used to account for missing data, which is indicated to be an appropriate choice of method when the data are MCAR (Asparouhov & Muthén, 2010).

The EFA was first conducted separately for each of the four lower-order factors, with the aim to refine the item pool and ensure one-dimensionality of each of the factors. Given that one higher-order and four lower-order factors were already expected based on the conceptual framework, retaining a one-factor solution for each of the domains was an important consideration not to add complexity to the model. More complex structures require more statistical power, and obtaining a larger sample size for the current study to increase statistical power was not possible due to resource and time limitations.

For each domain, a polychoric correlation matrix was created to account for the ordinal nature of the data. The correlation coefficients were extracted and used throughout the subsequent analyses. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO-MSA; Kaiser, 1974) and Bartlett's Test of Sphericity (Bartlett, 1951) were conducted to evaluate the suitability of the data for factor analysis. Consequently, an EFA was conducted for each of the domains, utilising an iterative approach. A one-factor model was first specified, and the factor loadings were inspected. All the items with factor loadings of $< .40$ were removed from the pool, following which EFA was conducted again on the remaining items. The process was repeated until no items with factor loadings of $< .40$ were present.

Next, parallel analysis was conducted for each of the domains to compare actual and simulated eigenvalues and inform the decisions regarding factor retention. Appropriate EFA models were specified, with the number of factors indicated by the parallel analysis. However, in an effort for the scale to remain consistent with the conceptual framework, decisions about factor retention were also guided by careful consideration of Lynch's (2018) model, with a primary focus of achieving a solution

with one theoretically meaningful factor per domain. Any changes made to the item pool against the initial parallel analysis were validated by additional parallel analyses and analyses of the consequent EFA models. Finally, parallel analyses and EFA were also conducted on the entire item pool, based on which the structure of the global scale was examined, and the item pool further refined.

6.5.1 The low flexible control (LFC) domain.

The KMO-MSA test indicated that the data was suitable for factor analysis, with an overall MSA value of .87 and item-MSA values ranging from .83 to .92. The values exceeded the recommended cut-off of .70 (Hair et al., 2019; Williams et al., 2010). The Bartlett's Test of Sphericity was significant ($\chi^2(120) = 500.91, p < .001$), also denoting factor analysis as a suitable analysis method (Hair et al., 2019; Williams et al., 2010).

A one-factor EFA model was specified using a polychoric correlation matrix. The output indicated three items with factor loadings of $< .40$, which were subsequently deleted. Then, a second one-factor EFA model was specified. The results indicated one additional item with a factor loading of $< .40$. As such, the item was deleted, following which a third one-factor EFA model was specified. The results highlighted one additional item with a factor loading below the threshold, which was also deleted from the item pool. Finally, a fourth one-factor model was specified, which indicated that all the remaining items ($N = 11$) had factor loadings above the selected threshold. For factor loadings of the rejected items, see Table 6.3. A parallel analysis was conducted with the remaining items. The results indicated that one factor should be retained (eigenvalue = 4.15, simulated eigenvalue = .44; see Figure 6.1 for a scree plot).

Table 6.3: Items in the LFC domain that were rejected based on consequent EFA models.

Item	Factor loading
1st EFA model	
In day-to-day situations, I rehearse what I want to say over and over again.	.34
I persist at tasks even if they cause me distress.	.33

I must do what <i>I</i> believe is right.	.31
2nd EFA model	
Being correct is more important to me than it is to other people.	.39
3rd EFA model	
I get annoyed when people don't adhere to social norms and standards.	.38

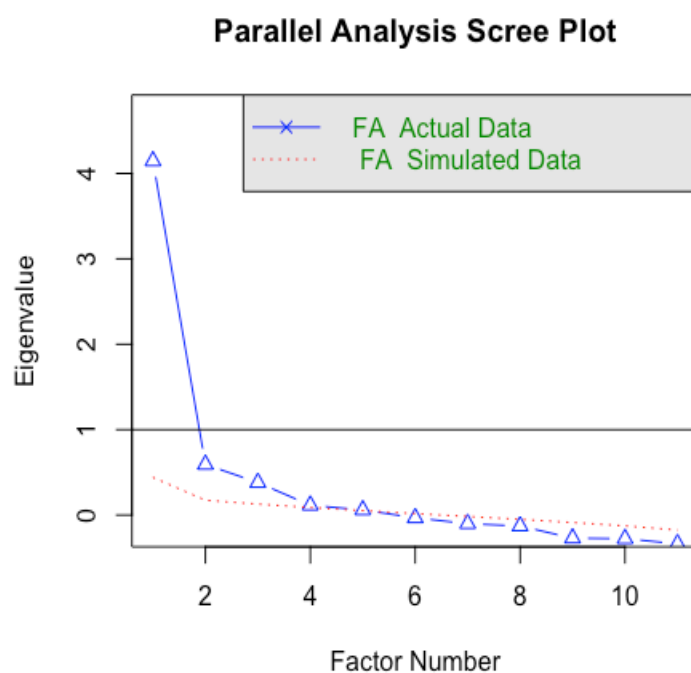


Figure 6.1: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the LFC domain.

Consistent with the parallel analysis results, a unidimensional, 11-item model of the LFC domain was established (see Table 6.4). The results of the EFA indicated that the model explained 37.69% of the variance.

Table 6.4: Items retained in the LFC domain following lower-order factor EFA.

Item	Factor loading
I like to be prepared for any possibility.	.63
I make detailed plans for how to do things.	.68
I feel the need to fix problems immediately.	.63

I dislike it when plans change unexpectedly.	.59
Structure and order in everyday life are important to me.	.79
I dislike it when my daily routine is interrupted.	.69
I like my life well-structured and predictable.	.75
Having self-control is more important to me than it is to most people.	.51
I pride myself in appearing to be in complete control of my behaviour and emotions.	.41
I cannot stop thinking about a problem until I find a solution.	.55
One should prioritise responsibilities over having fun.	.42

Note. All loadings were statistically significant at the 0.05 level.

6.5.2 The low receptivity and openness (LRO) domain.

All the items assigned to the LRO domain were selected ($N = 23$). The KMO-MSA (MSA index = .88, range: .70-.93) and the Bartlett's Test of Sphericity ($\chi^2(253) = 815.78, p < .001$) indicated that the data were suitable for factor analysis.

A one-factor EFA model was specified using a polychoric correlation matrix. The output indicated seven items with factor loadings of $< .40$, all of which were deleted. Then, a second one-factor EFA model was specified. The results indicated one additional item with a factor loading of $< .40$. As such, the item was deleted, following which a third one-factor EFA model was specified, which indicated that all the remaining items ($N = 15$) had factor loadings above the chosen threshold. For factor loadings of the rejected items, see Table 6.5.

Table 6.5: Items in the LRO domain that were rejected based on consequent EFA models.

Item	Factor loading
1st EFA model	
When I notice small discrepancies, I must fix them immediately.	.39
When someone gives me negative feedback, I just ignore it.	.03
I carefully consider potential consequences before taking a risk.	.29
*I welcome critical feedback even if I did not ask for it.	.36
*I often do things just for fun.	.37
*Life without risks is no fun.	.30
When someone tells me something I do not want to hear about myself, I explain to them why they are wrong about me.	.18
2nd EFA model	
*Unpredictability is exciting.	.32

Note. Reverse-scored items are indicated with an asterisk (*).

A parallel analysis was conducted with the remaining items. The results indicated that one factor should be retained (eigenvalue = 5.04, simulated eigenvalue = .42; see Figure 6.2 for a scree plot).

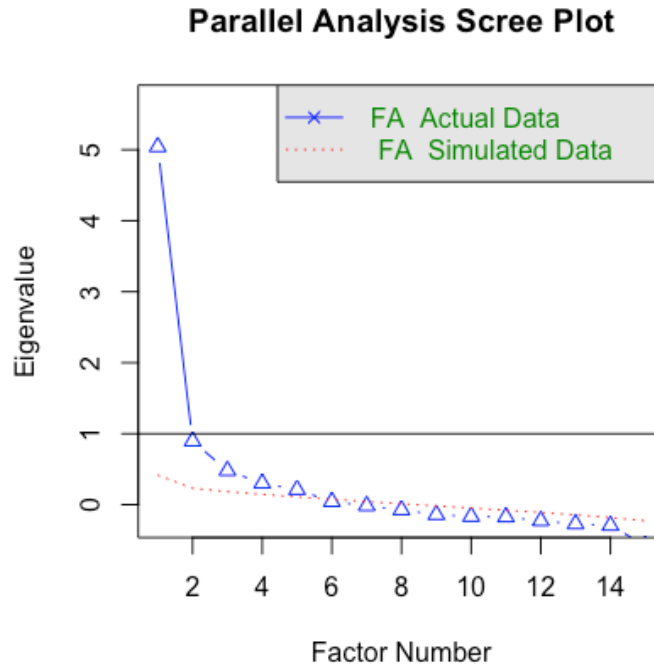


Figure 6.2: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the LRO domain.

Guided by the parallel analysis results, a one-factor, 15-item model of the LRO domain was retained. The results of the EFA indicated that all the items had factor loadings $\geq .40$, and that the model explained 33.61% of the variance. For a list of LRO items retained within the model, see Table 6.6.

Table 6.6: Items retained in the LRO domain following lower-order factor EFA.

Item	Factor loading
I would only volunteer to answer a question if I was confident I knew the correct answer.	.52
I avoid situations in which I could be criticised.	.70
*I am willing to spontaneously try new things.	.43
I avoid situations in which I could be seen as weak or incompetent.	.65
I am typically on high alert.	.61
The world is a scary place.	.51
When I make a mistake in front of others, I feel more embarrassed than most people would.	.73
Avoiding negative consequences is a better motivator than gaining something positive.	.56
I often see a threat where others may not.	.57
I dislike uncertainty more than most people.	.63
*I am happy to get out of my comfort zone and try something new.	.44
I avoid being the centre of attention.	.44
Embarrassing myself in front of others is one of my worst fears.	.68
When someone gives me a compliment, I ask myself why they are complimenting me.	.58

If people give me positive feedback, it is because they want to manipulate me.	.52
--	-----

Note. All loadings were statistically significant at the 0.05 level. * Reverse-scored items are indicated with an asterisk (*).

6.5.3 The low social connectedness and intimacy with others (LSC) domain.

All the items assigned to the LSC domain were selected ($N = 27$). The KMO-MSA (MSA index = .93, range: .78-.96) and the Bartlett's Test of Sphericity ($\chi^2(351) = 1268.16, p < .001$) indicated that the data were suitable for factor analysis.

A one-factor EFA model was first specified using a polychoric correlation matrix. The results revealed three items with factor loadings of $< .40$, which were subsequently deleted (see Table 6.7). Then, a new one-factor EFA model was specified, which indicated that all the remaining items ($N = 24$) had factor loadings above the chosen threshold. As such, no more items were rejected at this stage. Next, parallel analysis was conducted with the remaining items. The results indicated that two factors should be retained (eigenvalues: Factor 1 = 9.16, Factor 2 = 1.61; simulated eigenvalues: Factor 1 = .45, Factor 2 = .33; see Figure 6.3 for a scree plot).

Table 6.7: Items in the LSC domain rejected based on the one-factor EFA model.

Item	Factor loadings
When conflict arises, I withdraw from the situation.	.38
I envy people who do better than me.	.27
If I were to cry, I would only do it on my own or in front of people that I am close to.	.28

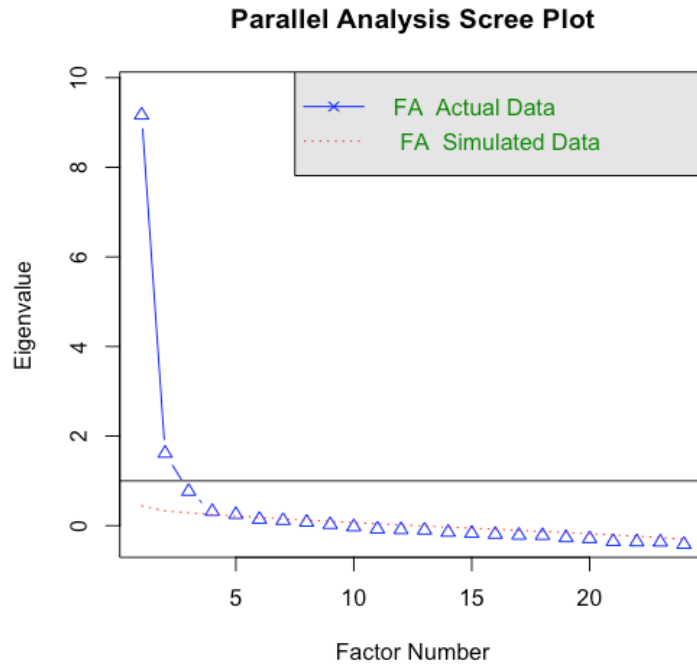


Figure 6.3: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the LSC domain.

Guided by the parallel analysis, a two-factor EFA model was specified, and the results indicated that two items had factor loadings of $<.40$ on both factors (see Table 6.8). The items were deleted, and parallel analysis was ran on the refined item pool, again indicating a two-factor solution (eigenvalues: Factor 1 = 8.57, Factor 2 = 1.60; simulated eigenvalues: Factor 1 = .43, Factor 2 = .31; see Figure 6.4 for a scree plot).

As per the results of the parallel analysis, a new two-factor EFA model was specified. The results indicated that all the items ($N = 22$) had factor loadings $\geq .40$ on one factor only, and therefore no more items were rejected at this stage. Within the two-factor model, Factor 1 was indicated to explain 26.81% of the variance, while Factor 2 was indicated to explain 20.43% of the variance, with an inter-factor correlation coefficient of .63 (95% CI = .59, .67).

Table 6.8: Items in the LSC domain rejected based on the two-factor EFA model.

Item	Factor 1 loading	Factor 2 loading
I dislike pointless social interactions.	.39	.33
I hold grudges for a long time.	.15	.31

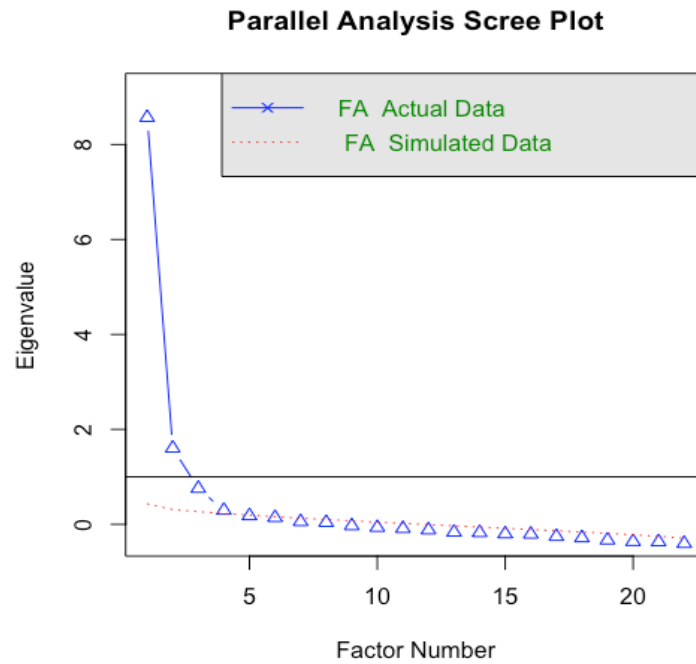


Figure 6.4: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the LSC domain following the rejection of two items with factor loadings <.40.

The items belonging to each of the two factors were inspected (see Table 6.9), and clear conceptual differences between the factors were observed. Factor 1 items ($N = 10$) pertained to sociability/gregariousness and extraversion/introversion, while Factor 2 items ($N = 12$) represented qualities related to social connectedness/alooness and empathic understanding of others. Both factors were also considered in the light of the conceptual framework. It appeared that as a standalone element, Factor 1 did not fit the behavioural manifestations of the LSC deficit well posited by Lynch (2018). Meanwhile, Factor 2 items closely aligned with the conceptual framework, with the items covering most aspects of the LSC deficit. As one of the purposes of establishing a conceptual framework in the early stages of the psychometric process is to establish boundaries of a construct, the researcher decided to only retain Factor 2 to stay within the established boundaries. The ten items that loaded on Factor 1 were, therefore, rejected. Following the removal of Factor 1 items, parallel analysis was conducted again to establish whether a one-factor solution was sufficient. Indeed, the results indicated that one factor should be retained (eigenvalue = 4.46, simulated eigenvalue = .43; see Figure 6.5 for a scree plot).

A final, one-factor EFA model was specified, and the results inspected. All the items had factor loadings $\geq .40$, and thus no additional items were rejected. A 12-item, one-factor solution was retained for the LSC domain. The output indicated that the model explained 37.20% of the variance. For factor loadings of items retained within the model, see Table 6.10.

Table 6.9: Factor loadings of items in the LSC domain indicated by a two-factor EFA.

Item	Factor 1 loading	Factor 2 loading
In social situations, I prefer it when people have clearly assigned roles.	.12	.40
I rarely feel deeply connected to other people.	.19	.57
I have few close friends, if any.	.40	.30
I feel that I am an outsider.	.33	.45
*Being around other people makes me feel energised.	.95	-.23
*I am a sociable person.	.85	-.04
I prefer being on my own rather than around other people.	.69	.16
I feel the need to be alone for a while after participating in social events.	.71	-.02
It takes me a long time to warm up to new people.	.46	.35
I dislike parties.	.71	.02
*I enjoy meeting new people.	.89	-.14
I often struggle to understand another person's perspective.	-.21	.68
*Being around other people helps me relax.	.93	-.23
My relationships with other people are rather shallow.	.10	.62
I behave more awkwardly in social situations than most people do.	.44	.30
I struggle to believe in real, unconditional love.	-.04	.52
When other people become upset, I don't know how to support them.	-.11	.75
I dislike being around people who are upset.	-.10	.65
I am different from other people.	-.08	.40
When I see another person crying, I find it difficult to understand why they would do this in public.	-.17	.67
I am distant from other people.	.39	.55
I am misunderstood by other people.	.17	.47

Note. Factor loadings $> .40$ are presented in bold. Reverse-scored items are indicated with an asterisk (*).

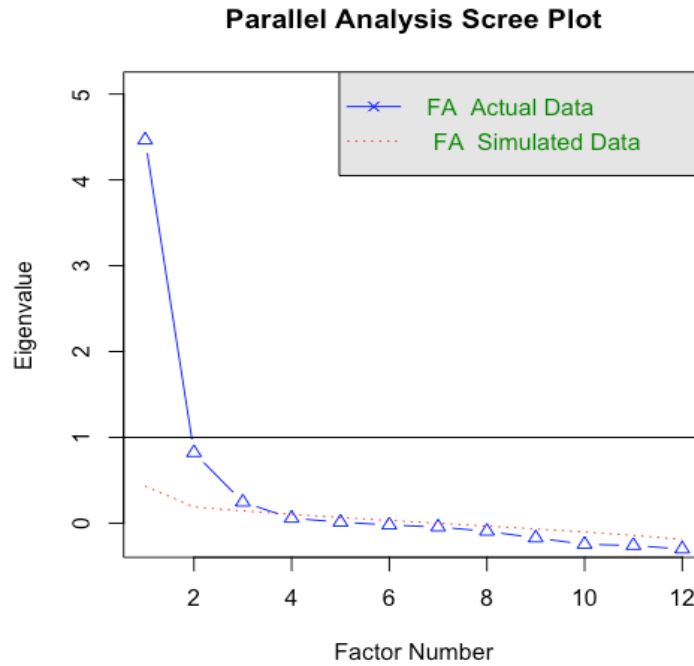


Figure 6.5: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the LSC domain after the rejection of Factor 1.

Table 6.10: Items retained in the LSC domain following lower-order factor EFA.

Item	Factor loading
In social situations, I prefer it when people have clearly assigned roles.	.46
I rarely feel deeply connected to other people.	.71
I feel that I am an outsider.	.70
I often struggle to understand another person's perspective.	.50
My relationships with other people are rather shallow.	.68
I struggle to believe in real, unconditional love.	.51
When other people become upset, I don't know how to support them.	.65
I dislike being around people who are upset.	.57
I am different from other people.	.46
When I see another person crying, I find it difficult to understand why they would do this in public.	.53
I am distant from other people.	.82
I am misunderstood by other people.	.60

Note. All loadings were statistically significant at the 0.05 level.

6.5.4 The pervasive inhibited emotional expression and low emotional awareness (PIE) domain.

All the items assigned to the PIE domain were selected ($N = 25$). The KMO-MSA (MSA index = .91, range: .87-.94) and the Bartlett's Test of Sphericity ($\chi^2(300) = 976.29, p < .001$) indicated that the data were suitable for factor analysis.

A one-factor EFA model was first specified using a polychoric correlation matrix. The results revealed five items with factor loadings of $< .40$, all which were subsequently rejected (see Table 6.11). Then, a new one-factor EFA model was specified, which indicated that all the remaining items ($N = 20$) had factor loadings $\geq .40$, meeting the threshold. As such, no additional items were deleted at this stage.

Table 6.11: Items in the PIE domain rejected based on the one-factor EFA model.

Item	Factor loading
My facial expressions don't always match how I really feel (for example, I laugh when I feel awkward, or I smile when I'm sad).	.37
I consciously put on facial expressions that I think are appropriate in a given situation.	.32
*I experience positive emotions strongly.	.34
It is difficult for me to stop worrying.	.04
To support somebody is to help them find a solution to the problem.	.25

Parallel analysis was conducted with the remaining items. The results indicated that two factors should be retained (eigenvalues: Factor 1 = 6.46, Factor 2 = 1.67; simulated eigenvalues: Factor 1 = .42, Factor 2 = .29; see Figure 6.6 for a scree plot).

In accordance with the results of parallel analysis, a new, two-factor EFA model was specified. The results revealed that 14 items loaded on Factor 1 and six items loaded on Factor 2. All the items had factor loadings $\geq .40$ on one factor only, and as such, no additional items were rejected at this stage. In the two-factor model, Factor 1 ($N = 14$) was indicated to explain 24.62% of the variance, while Factor 2 ($N = 6$) was indicated to explain 17.48% of the variance, with an inter-factor correlation coefficient of .54 (95% CI = .48, .59).

The items belonging to each of the two factors were inspected (see Table 6.12). Despite no multi-collinearity between the factors, after a careful consideration, no ambiguous conceptual differences between the factors were evident, with both containing items that consider both feeling and communicating emotions. For instance, while Factor 2 items appeared to have tap into more negative attitudes towards experiencing and expressing emotions (e.g., '*I avoid dwelling on my emotions.*') when compared to Factor 1, this was not unique, with Factor 1 also containing some items that could be interpreted in a negative way (e.g., '*Being emotional makes me feel like I'm a weak person.*').

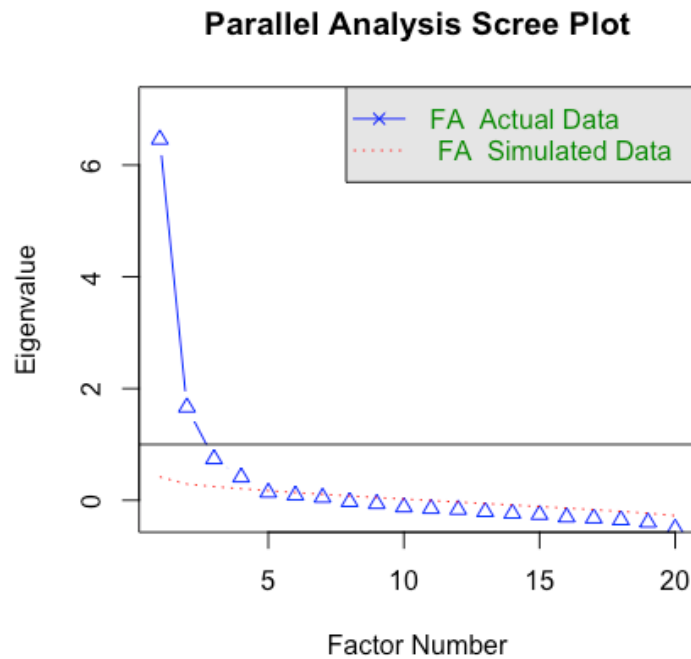


Figure 6.6: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the PIE domain.

Table 6.12: Factor loadings of items in the PIE domain indicated by a two-factor EFA.

Item	Factor 1 loading	Factor 2 loading
I rarely express my excitement to others.	.54	.12
I tend to dismiss my emotions.	.60	.12
Being emotional makes me feel like I'm a weak person.	.73	-.20
Other people feel emotions much more intensely than I do.	.12	.58
I don't let my emotions get to me.	-.26	.90
I choose reason over emotion.	.01	.63
*I am an open book to other people.	.52	-.05
I push through difficult situations without sharing my struggles with others.	.50	.23
I don't reveal my vulnerability.	.56	.25
I rarely complain about being emotionally hurt.	.08	.69
I rarely complain about being distressed.	.03	.68
Other people express emotions more openly than I do.	.52	.34
No matter what I feel on the inside, I make sure I seem fine on the outside.	.43	.19
Thinking about myself crying makes me feel uneasy.	.64	-.10
I tend to bottle up my feelings.	.85	-.16
I tell others that I'm fine, even if I'm not.	.70	-.18
I avoid dwelling on my emotions.	-.19	.79

*When something bad happens, talking to other people about it helps me a lot.	.42	.08
People have told me that I'm difficult to read.	.62	-.01
*People see me as a warm and affectionate person.	.40	.06

Note. Factor loadings > .40 are presented in bold. * Reverse-scored items are indicated with an asterisk (*).

Both factors were also considered in view of the conceptual framework. Items that loaded on both factors were directly relevant to the conceptual framework. However, it appeared that as a standalone element, Factor 1 items covered many more of the typical behavioural manifestations of the PIE deficit (Lynch, 2018). Moreover, bearing in mind the intended meaning of the items in the content of maladaptive overcontrol (and specifically the PIE deficit), the researcher felt that most Factor 2 items did not draw upon any traits or behaviours that were not already represented by Factor 1 items alone. This was except for two Factor 2 items, one of which denoted strong detachment from one's emotional states (*I don't let my emotions get to me.*), and one of which represented a preference towards logical reasoning over emotion-driven decision-making (*I choose reason over emotion.*)

Considering the difficulty in interpreting the factors as distinctly different in the light of the conceptual model, relatively little unique content being lost if Factor 2 was rejected, and the goal to keep the measure structure as simple as possible for the purpose of both interpretability and statistical power, a decision was made to only retain Factor 1. As such, the six items belonging to Factor 2 items were rejected.

Following the removal of Factor 2 items, parallel analysis was conducted again, the results of which indicated that a one-factor solution was sufficient (eigenvalue = 5.15, simulated eigenvalue = .42; see Figure 6.7 for a scree plot). As such, a one-factor EFA model was specified. The results indicated that all the items had factor loadings $\geq .40$. As such, no additional items were rejected, and a 14-item, one-factor solution was retained for the PIE domain. The model explained 36.81% of the variance. For a list of PIE items retained within the model, see Table 6.13.

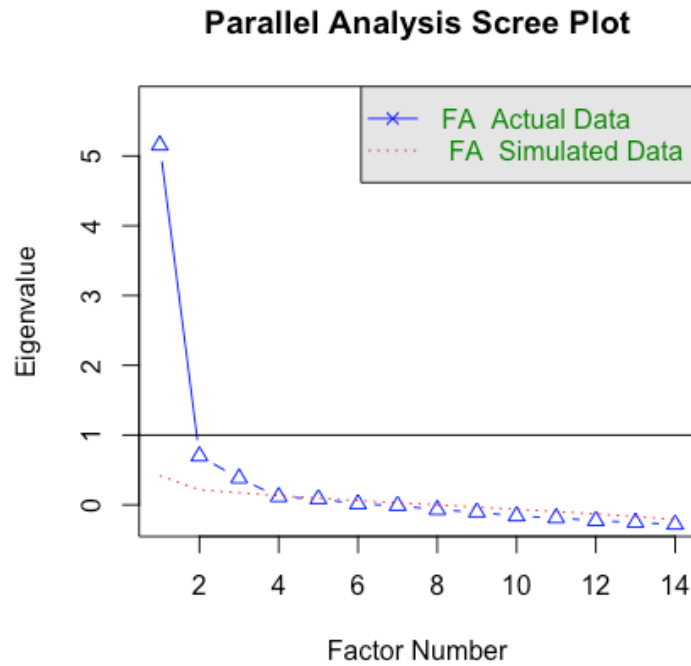


Figure 6.6: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the PIE domain following removal of Factor 2 items.

Table 6.13: Items retained in the PIE domain following lower-order factor EFA.

Item	Factor loading
I rarely express my excitement to others.	.62
I tend to dismiss my emotions.	.67
Being emotional makes me feel like I'm a weak person.	.60
*I am an open book to other people.	.49
I push through difficult situations without sharing my struggles with others.	.64
I don't reveal my vulnerability.	.70
Other people express emotions more openly than I do.	.71
No matter what I feel on the inside, I make sure I seem fine on the outside.	.54
Thinking about myself crying makes me feel uneasy.	.57
I tend to bottle up my feelings.	.75
I tell others that I'm fine, even if I'm not.	.59
*When something bad happens, talking to other people about it helps me a lot.	.47
People have told me that I'm difficult to read.	.62
*People see me as a warm and affectionate person.	.44

Note. All loadings were statistically significant at the 0.05 level. * Reverse-scored items are indicated with an asterisk (*).

6.5.4 The global EFA model.

Following lower-order EFA analysis for each of the domains, a higher-order EFA including the retained items across the four domains was conducted. This was to verify whether the model was best explained as a four-factor model, and if the items best fit in the domain that they had been assigned to. First, a parallel analysis was conducted to test dimensionality of the model. The results indicated that not four, but five factors

should be retained (eigenvalues: 13.13, 3.96, 2.02, 1.67, 1.46; simulated eigenvalues, respectively: .66, .57, .53, .49, .46; see Figure 6.7 for a scree plot).

A five-factor EFA model was specified. The factor loadings of items were inspected, and it appeared that Factor 1 reflected the LRO domain (variance explained: 10.98%), Factor 2 reflected the LSC domain (variance explained: 10.82%), Factor 3 reflected the LFC domain (variance explained: 9.56%), and Factor 4 reflected the PIE domain (variance explained: 8.56%). Factor 5 (variance explained: 5.18%) only contained two items, both of which were reverse-scored items originally assigned to the LRO domain and tapped into the respondent's attitudes on engaging in spontaneous and novel activities.

While Factor 5 appeared relevant to the LRO domain's definition within the conceptual framework, psychometric literature indicates that factors with less than three items are generally weak and unstable and should not be retained (Costello & Osborne, 2005). As such, the two items that comprised Factor 5 were rejected. Additionally, 5 items that had factor loadings $> .40$ on domains different than they were originally assigned to, 3 items that had factor loadings $> .40$ on more than one domain, and 10 items that did not have factor loadings $> .40$ on any of the domains were also rejected. For factor loadings of each of the items, please refer to Table 6.14.

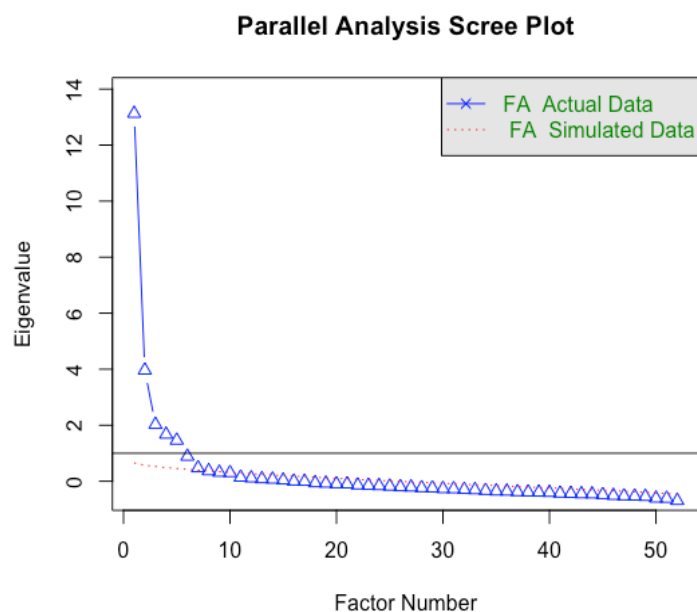


Figure 6.7: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the first global EFA model.

Table 6.14: Factor loadings of items indicated by a five-factor global EFA.

Item	Domain	F1 (LRO)	F2 (LSC)	F3 (LFC)	F4 (PIE)	F5
I like to be prepared for any possibility.	LFC	.09	-.23	.63	.16	-.09
I make detailed plans for how to do things.	LFC	.10	-.08	.67	-.07	-.02
I feel the need to fix problems immediately.	LFC	.14	-.10	.61	.06	-.17
I dislike it when plans change unexpectedly.	LFC	.17	.07	.45	-.17	.32
Structure and order in everyday life are important to me.	LFC	-.14	-.03	.82	-.03	.15
I dislike it when my daily routine is interrupted.	LFC	.05	.14	.59	-.13	.27
I like my life well-structured and predictable.	LFC	-.12	.08	.76	-.12	.31
Having self-control is more important to me than it is to most people.	LFC	-.11	.13	.51	.29	-.14
I pride myself in appearing to be in complete control of my behaviour and emotions.	LFC	-.28	-.03	.46	.57	-.12
I cannot stop thinking about a problem until I find a solution.	LFC	.32	-.08	.46	.05	-.20
One should prioritise responsibilities over having fun.	LFC	-.08	.06	.42	.14	-.01
I would only volunteer to answer a question if I was confident I knew the correct answer.	LRO	.22	-.11	.14	.21	.28
I avoid situations in which I could be criticised.	LRO	.49	.07	.02	.01	.23
*I am willing to spontaneously try new things.	LRO	-.03	.06	.08	-.18	.79
I avoid situations in which I could be seen as weak or incompetent.	LRO	.38	.10	.14	.09	.16
I am typically on high alert.	LRO	.63	-.07	.27	-.04	-.06
The world is a scary place.	LRO	.57	-.08	.08	-.05	.03
When I make a mistake in front of others, I feel more embarrassed than most people would.	LRO	.68	-.16	.00	.07	.25
Avoiding negative consequences is a better motivator than gaining something positive.	LRO	.25	.17	.16	.01	.18
I often see a threat where others may not.	LRO	.54	.01	.27	.00	-.10
I dislike uncertainty more than most people.	LRO	.29	.09	.38	-.17	.33
*I am happy to get out of my comfort zone and try something new.	LRO	.06	.01	-.04	-.14	.81
I avoid being the centre of attention.	LRO	.13	.03	.01	.20	.34
Embarrassing myself in front of others is one of my worst fears.	LRO	.60	-.29	-.01	.19	.29
When someone gives me a compliment, I ask myself why they are complimenting me.	LRO	.67	.06	-.21	.10	.04
If people give me positive feedback, it is because they want to manipulate me.	LRO	.41	.42	-.04	-.01	-.11
In social situations, I prefer it when people have clearly assigned roles.	LSC	.15	.22	.46	-.04	.09
I rarely feel deeply connected to other people.	LSC	.11	.60	-.06	.07	.03
I feel that I am an outsider.	LSC	.56	.37	-.15	-.04	-.01
I often struggle to understand another person's perspective.	LSC	.05	.56	.04	-.17	-.01
My relationships with other people are rather shallow.	LSC	.16	.66	-.02	-.10	-.06
I struggle to believe in real, unconditional love.	LSC	.21	.43	-.04	.04	-.12

When other people become upset, I don't know how to support them.	LSC	-.03	.68	.02	-.01	.08
I dislike being around people who are upset.	LSC	-.14	.70	.17	-.05	-.05
I am different from other people.	LSC	.42	.26	.07	-.08	-.27
When I see another person crying, I find it difficult to understand why they would do this in public.	LSC	-.22	.58	.10	.19	.05
I am distant from other people.	LSC	.28	.62	-.09	.04	.08
I am misunderstood by other people.	LSC	.60	.32	-.01	-.11	-.14
I rarely express my excitement to others.	PIE	-.05	.53	-.12	.27	.08
I tend to dismiss my emotions.	PIE	.13	.19	-.09	.51	-.07
Being emotional makes me feel like I'm a weak person.	PIE	.29	.26	.03	.27	-.02
*I am an open book to other people.	PIE	-.03	.05	-.15	.41	.27
I push through difficult situations without sharing my struggles with others.	PIE	.01	.02	.09	.69	-.17
I don't reveal my vulnerability.	PIE	-.11	.12	.07	.70	.03
Other people express emotions more openly than I do.	PIE	-.20	.35	.06	.56	.07
No matter what I feel on the inside, I make sure I seem fine on the outside.	PIE	.08	-.13	.09	.69	-.16
Thinking about myself crying makes me feel uneasy.	PIE	.23	.25	.09	.28	-.03
I tend to bottle up my feelings.	PIE	.31	.05	-.11	.57	-.02
I tell others that I'm fine, even if I'm not.	PIE	.47	-.26	-.09	.60	-.03
*When something bad happens, talking to other people about it helps me a lot.	PIE	-.07	.36	-.15	.24	.10
People have told me that I'm difficult to read.	PIE	.14	.32	-.06	.33	-.06
*People see me as a warm and affectionate person.	PIE	-.14	.59	-.08	.05	.11

Note. Factor loadings > .40 are presented in bold. Items rejected based on the presented factor loadings are shaded in grey. *Reverse-scored items are indicated with an asterisk (*).

Subsequently, an additional parallel analysis was conducted to inspect the dimensionality of the entire scale again. The results indicated that a four-factor solution should be retained (eigenvalues: 7.85, 3.09, 1.30, 1.17; simulated eigenvalues, respectively: .51, .41, .37, .33; see Figure 6.8 for a scree plot).

A four-factor EFA model was subsequently specified. The results indicated that one additional item (*I struggle to believe in real, unconditional love.*), originally assigned to the LSC domain, did not have a factor loading > .40 on either of the domains and was therefore removed. The remaining items all had a factor loading > .40 on one domain only, and for all the items, it was the domain that aligned with the originally assigned domain.

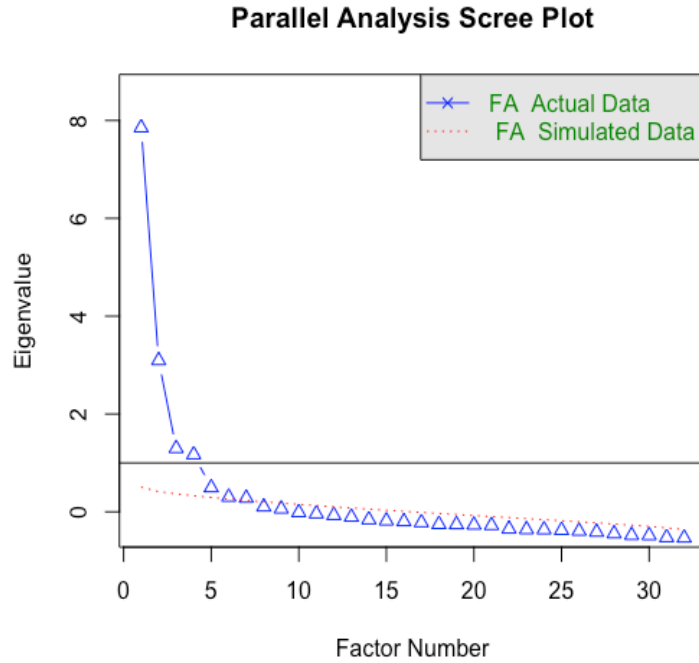


Figure 6.8: Parallel analysis scree plot depicting actual and simulated eigenvalues of factors for the second global EFA model.

All the remaining items had a factor loading $> .40$ on one domain only, and for all the items, it was the domain that aligned with the originally assigned domain ($N_{LFC} = 10$; $N_{LRO} = 7$; $N_{LSC} = 7$; $N_{PIE} = 7$). As such, no additional items were deleted. For a full list of items retained within the four-factor model following global EFA, refer to Table 6.15.

Table 6.15: Factor loading of items in a final, four-factor global EFA model.

Item	Domain	F1 (LFC)	F2 (LSC)	F3 (LRO)	F4 (PIE)
I like to be prepared for any possibility.	LFC	.66	-.27	.03	.18
I make detailed plans for how to do things.	LFC	.69	-.05	.06	-.09
I feel the need to fix problems immediately.	LFC	.62	-.12	.06	.03
I dislike it when plans change unexpectedly.	LFC	.48	.21	.22	-.17
Structure and order in everyday life are important to me.	LFC	.83	.09	-.10	-.11
I dislike it when my daily routine is interrupted.	LFC	.61	.27	.10	-.15
I like my life well-structured and predictable.	LFC	.76	.20	-.06	-.14
Having self-control is more important to me than it is to most people.	LFC	.46	.05	-.12	.29
I cannot stop thinking about a problem until I find a solution.	LFC	.48	-.13	.23	.07
One should prioritise responsibilities over having fun.	LFC	.42	.02	-.11	.16
I avoid situations in which I could be criticised.	LRO	.01	.19	.55	-.03

I am typically on high alert.	LRO	.28	-.06	.61	-.04
The world is a scary place.	LRO	.10	-.03	.55	-.05
When I make a mistake in front of others, I feel more embarrassed than most people would.	LRO	-.01	.01	.76	-.02
I often see a threat where others may not.	LRO	.29	-.02	.45	.04
Embarrassing myself in front of others is one of my worst fears.	LRO	-.03	-.10	.71	.08
When someone gives me a compliment, I ask myself why they are complimenting me.	LRO	-.17	.06	.61	.14
I rarely feel deeply connected to other people.	LSC	-.07	.52	.12	.19
I often struggle to understand another person's perspective.	LSC	.01	.62	.03	-.16
My relationships with other people are rather shallow.	LSC	-.04	.56	.14	.03
When other people become upset, I don't know how to support them.	LSC	-.03	.78	.02	-.01
I dislike being around people who are upset.	LSC	.13	.67	-.15	.01
When I see another person crying, I find it difficult to understand why they would do this in public.	LSC	.03	.58	-.17	.23
I am distant from other people.	LSC	-.06	.51	.26	.20
I tend to dismiss my emotions.	PIE	-.11	.17	.12	.50
*I am an open book to other people.	PIE	-.10	.03	.05	.43
I push through difficult situations without sharing my struggles with others.	PIE	.10	-.08	-.04	.71
I don't reveal my vulnerability.	PIE	.07	.03	-.06	.74
Other people express emotions more openly than I do.	PIE	.05	.31	-.17	.62
No matter what I feel on the inside, I make sure I seem fine on the outside.	PIE	.07	-.13	.05	.63
I tend to bottle up my feelings.	PIE	-.10	.02	.29	.61

Note. All loadings were statistically significant at the 0.05 level. * Reverse-scored items are indicated with an asterisk (*).

Lastly, inter-factor correlations were inspected (see Table 6.16). All the correlations between the factors were positive. The weakest correlations were found between the LFC domain and the LSC and PIE domains, while the strongest correlations were found between the LFC domain and the LRO domain and the LRO domain and the PIE domain.

Table 6.16

Factor correlations and confidence intervals in the final, four-factor EFA model.

	LFC	LRO	LSC	PIE
	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
LFC	-	.51 (.20, .66)	.23 (.11, .58)	.24 (.07, .62)
LRO	.51 (.20, .66)	-	.46 (.24, .59)	.50 (.26, .65)

6.6 Confirmatory Factor Analysis (CFA): Data analysis and results.

The weighted least squares with means and variance adjusted (*wlsmv*) estimator was used throughout due to its suitability for use with Likert-scale data (the sample size in the current study met the recommended size of > 200 ; Beauducel & Herzberg, 2006; Li, 2016). The ordinal nature of the data was explicitly specified within the model. Pairwise deletion was applied to deal with missing values, which was indicated to be a suitable method when the data are MCAR (Asparouhov & Muthén, 2010).

In accordance with the conceptual framework, an initial higher-order CFA model was specified with overcontrol as the higher-order factor and, in line with the EFA results, a 10-item LFC, a 7-item LRO, a 7-item LSC, and a 7-item PIE as the four lower-order factors. The initial inspection of model fit indices ($\chi^2(465) = 2588.14$, $p < .001$, CFI = .928, TLI = .922, RMSEA = .092 (90% CI [.090, .097], SRMR = .085) did not indicate an acceptable fit of the model according to the criteria proposed by Hair and colleagues (2019). As such, steps were taken to improve model fit, with caution not to undermine the theoretical validity of the instrument.

Standardised factor loadings were first inspected, with the intention to remove all the items for which the value was $< .40$, consistently with the cut-off used in EFA. The results of the CFA indicated that one item (*One should prioritise responsibilities over having fun.*), from the LFC domain, had a factor loading that did not meet the cut-off ($= .39$). As such, the item was deleted. A new CFA model was then specified, and factor loadings were inspected again. The output indicated that there were no additional items with factor loadings below the threshold. The results returned the following goodness of fit values: $\chi^2(401) = 2514.57$, $p < .001$, CFI = .928, TLI = .922, RMSEA = .094 (90% CI: [.088, .096]), SRMR = .083.

Modification indices were subsequently examined. Two indices with an equal highest initial value of 346.63 indicated that there was unaccounted covariance between the LFC and the LRO domains, and the LSC and the PIE domains. This suggested that model fit could be improved if direct paths between the respective domains was specified within the model. However, the Neurobiosocial Theory for Disorders of Overcontrol (Lynch, 2018) is a novel theory that is yet to be thoroughly empirically tested – meaning that the relationships between specific deficits identified

within the theoretical model are unknown. Moreover, specifying the direct paths between the domains introduces additional complexity to the already complex model, creating potential issues with statistical power and overfitting. Seeking a compromise between data-driven model fit improvement and maintaining theoretical integrity, no direct paths between domains were specified. Instead, a decision was made to attempt to improve model fit via data-driven item reduction.

A stepwise process was applied to refine the item list and improve model fit. The next highest modification index was inspected to identify a problematic item that was indicated to load on a domain different than the one initially specified or loaded directly on the higher-order factor. The item was rejected, and a new CFA model was specified, and modification indices were inspected again. Based on the values of the indices, four more items were rejected:

- The item '*Having self-control is more important to me than it is to most people.*' from the LFC domain was removed. Modification index with a value of 178.31 indicated that the item had a stronger relationship with the PIE domain. Other modification indices involving this item additionally indicated a strong loading directly onto the higher order factor (= 157.35) and a cross-loading with both the LSC domain (= 116.95) and the LRO domain (= 89.36). Fit indices post-removal: $\chi^2(373) = 2260.30$, $p < .001$, CFI = .932, TLI = .926, RMSEA = .092 (90% CI [.088, .096]), SRMR = .083.
- The item '*I am typically on high alert.*' From the LRO domain was removed. Modification index with a value of 141.75 indicated that the item had a stronger relationship with the LFC domain. Other modification indices involving this item additionally indicated high collinearity with several items from the LFC domain, with the strongest relationship with the item *I cannot stop thinking about a problem until I find a solution.* (= 61.50). Fit indices post-removal: $\chi^2(346) = 1902.55$, $p < .001$, CFI = .941, TLI = .936, RMSEA = .087 (90% CI [.083, .090]), SRMR = .079.
- The item '*I dislike it when plans change unexpectedly.*' from the LFC domain was removed. Modification index with a value of 124.87 indicated that the item had a stronger relationship with the LRO domain. Other

modification indices involving this item additionally indicated a strong loading directly onto the higher order factor (= 66.09) and a cross-loading with both the LSC domain (= 52.33) and the PIE domain (= 26.07). Fit indices post-removal: $\chi^2(320) = 1641.70, p < .001$, CFI = .945, TLI = .940, RMSEA = .083 (90% CI [.079, .087]), SRMR = .077.

- The item '*I dislike it when my daily routine is interrupted.*' from the LFC domain) was removed. Modification index with a value of 180.66 indicated that the item had a stronger relationship with the LRO domain. Other modification indices involving this item additionally indicated a strong loading directly onto the higher order factor (= 135.82) and a cross-loading with both the LSC domain (= 122.65) and the PIE domain (= 80.60). Fit indices post-removal: $\chi^2(295) = 1389.44, p < .001$, CFI = .950; TLI = .945; RMSEA = .079 (90% CI [.075, .083]), SRMR = .074.

As reported above, after removal of the above items, the χ^2 test remained significant as predicted by Hair and colleagues (2019), and all other fit indices but the RMSEA met the goodness-of-fit criteria endorsed by the authors. Considering the authors' direction for caution and flexibility when using the cut-offs so that the scale's theoretical validity is not compromised, the fact that this is the first attempt to develop a measure based on Lynch's (2018) model, as well as the large sample size used and a relatively complex model applied in the current study, the decision was made not to make any further changes to the item pool.

A final higher-order CFA model was specified with the remaining items, with overcontrol as the higher-order factor, and the LFC (6 items), the LRO (6 items), the LSC (7 items), and the PIE (7 items) domains as the four lower-order factors. The standardised factor loadings were inspected once more, with all the items meeting the $\geq .40$ threshold. Standardised loadings of each domain on the higher order factor were also inspected, indicating a loading of .65 for the LRO domain, .84 for the LSC domain, and .79 for the PIE domain. However, the .40 threshold was not met for the LFC domain (= .37). Additionally, high levels of unaccounted covariance were also indicated between the LFC domain and the LSC domain and the LRO domain and the PIE domain (modification index value = 91.65 for both). For the list of retained items and their standardised loadings, please see Table 6.17.

Table 6.17: Standardised loadings of items retained following the refinement of the item pool based on the higher- order CFA results.

Item	Loading
LFC	
I like to be prepared for any possibility.	.50
I make detailed plans for how to do things.	.54
I feel the need to fix problems immediately.	.52
Structure and order in everyday life are important to me.	.90
I like my life well-structured and predictable.	.72
I cannot stop thinking about a problem until I find a solution.	.65
LRO	
I avoid situations in which I could be criticised.	.62
The world is a scary place.	.48
When I make a mistake in front of others, I feel more embarrassed than most people would.	.81
I often see a threat where others may not.	.55
Embarrassing myself in front of others is one of my worst fears.	.76
When someone gives me a compliment, I ask myself why they are complimenting me.	.62
LSC	
I rarely feel deeply connected to other people.	.73
I often struggle to understand another person's perspective.	.56
My relationships with other people are rather shallow.	.63
When other people become upset, I don't know how to support them.	.66
I dislike being around people who are upset.	.54
When I see another person crying, I find it difficult to understand why they would do this in public.	.59
I am distant from other people.	.82
PIE	
I tend to dismiss my emotions.	.61
*I am an open book to other people.	.53
I push through difficult situations without sharing my struggles with others.	.63
I don't reveal my vulnerability.	.65
Other people express emotions more openly than I do.	.77
No matter what I feel on the inside, I make sure I seem fine on the outside.	.58
I tend to bottle up my feelings.	.81

Note. *Reverse-scored items are indicated with an asterisk (*).

Following model refinement based on the CFA, alpha and omega values were calculated to assess the internal reliability of each of the subscales and the composite scale. Values of the ordinal alpha coefficient indicated that all four subscales reached an acceptable internal consistency threshold of $> .70$ (LFC = .80, LRO = .79, LSC = .83, PIE = .84). Omega values for each of the four subscales indicated that all of them reached an acceptable level of composite reliability of $> .70$ (LFC = .78, LRO = .80, LSC = .82, PIE = .82). Similarly, the higher-order omega value of .80 indicated an acceptable level of composite reliability of the higher-order factor of maladaptive overcontrol OAQ.

Of note, inter-factor correlations were not stated due to common variance in the specified higher-order model assumed to be explained by the higher-order factor of overcontrol. However, the final inspection of modification indices revealed that the recommendation to specify direct paths between the LFC and the LRO domain and the LSC and the PIE domain to account for covariance remained at the top of the list, albeit the value of modification index decreased to 171.55 for both.

6.7 Discussion

This chapter presented two studies conducted to develop the construct validity of the OAQ and validate its internal consistency. In the first study, EFA was used to (1) ensure that only the items most relevant to the theoretical framework are retained within the item pool, and (2) examine and refine the internal structure of the scale to ensure its consistency with the conceptual framework. In the second study, CFA was used to cross-validate the findings regarding the dimensionality of the scale, and further refine the pool as indicated by the factor loadings and modification indices. The following paragraphs provide an overview of the results and consider the strengths and limitations of the two studies.

6.7.1 Outcomes of the studies.

The EFA was initially conducted on each of the four theorised domains of maladaptive overcontrol separately. This was to ensure that only the most appropriate items are selected to measure each of the deficits, and that each domain is unidimensional. Each dimension being comprising only of one factor was crucial, in an effort not to introduce additional complexity to the conceptual model established based on Lynch's (2018) Neurobiosocial Theory for Disorders of Overcontrol.

After the initial rejection of items with factor loadings $< .40$ for each domain, parallel analysis indicated that the LFC domain and the LRO domain were, indeed, unidimensional. Both for the LSC domain and the PIE domain, however, the parallel analysis initially presented a two-factor solution. Upon examining the contents of the two factors comprising the LSC domain, the researcher noted that one of the factors was more representative of introversion/extraversion, sociability, and gregariousness of the respondents. While these constructs are likely related to maladaptive overcontrol, they are not explicitly present within the conceptual definition of the

construct. In turn, the other identified factor appeared to represent feelings of social disconnectedness, isolation, and difficulty empathising with others – attributes key to the definition of the LSC domain applied within this thesis. As such, only the latter factor was retained, with a subsequent parallel analysis supporting a one-factor solution. In the case of the PIE domain, there was no clear conceptual difference between the items comprising the identified factors. One of the factors, however, included a larger number of items, which captured a wider range of characteristics attributed to the PIE deficit. As such, the factor with a smaller number of items was rejected, with the later parallel analysis supporting the unidimensional structure of the domain.

Once a unidimensional solution was achieved for all deficits, a global EFA was conducted to further refine the item pool based on potential cross-loadings and ensure the structure fit the hypothesised conceptual framework – i.e., that no additional factors unexplained by the conceptual framework were present. The initial parallel analysis indicated the presence of five, rather than four, meaningful factors. The fifth, unexpected factor comprised only two items, both of which were originally assigned to the LRO domain, reverse-scored, and related to the respondents' openness to new experiences. It was an unexpected result for the items to load on a separate domain, as their content is evidently related to the LRO domain. A potential explanation for this is the reverse-scored items' tendency to load on unintended, separate factors (Dalal & Carter, 2014). Because retaining factors consisting of only two to three items is not recommended (Hair et al., 2019), the decision was made to reject the factor. Moreover, all the items that loaded on the domain different to what was expected based on the conceptual framework, returned cross-loadings, or did not load on either of the factors were also rejected. Parallel analysis was conducted again which, consistent with the conceptual framework, supported a four-factor structure.

The EFA study led to a rejection of 60 items (39 items at the subscale-only stage; $N_{LFC} = 5$, $N_{LRO} = 8$, $N_{LSC} = 15$, $N_{PIE} = 11$; 21 items at the global stage; $N_{LFC} = 1$, $N_{LRO} = 8$, $N_{LSC} = 5$, $N_{PIE} = 7$). As such, data pertaining to the rejected items were removed from the CFA dataset. A higher-order CFA model was then specified, with maladaptive overcontrol as the higher-order factor and the four EFA factors as the

lower-order factors, with each representing a theoretical domain (i.e., a core deficit) of maladaptive overcontrol posited by Lynch (2018).

Initial results of the CFA indicated the model fit to be less-than-satisfactory. Based on standardised loadings and modification indices five additional items were deleted (one item based on factor loadings and four items based on the value of the modification indices – $N_{LFC} = 4$, $N_{LRO} = 1$). After these changes were applied, only one fit index value remained below the expected goodness-of-fit threshold. Despite one of the indices not meeting the threshold, the researcher decided not to reject any additional items in pursuit of better model fit. This was for two reasons. Firstly, the structural model of maladaptive overcontrol within this thesis is relatively complex, and complexity is bound to inflate the values of the modification indices. This is because modification indices are calculated based on a change in chi-square value ($MI = \Delta\chi^2$), with the chi-square distribution being dependent on degrees of freedom (the difference between the number of observed variables p and parameters estimated in the model q : $df = \frac{1}{2} \times (p \times (p + 1) - q)$). More complex structural models, including the one in the current study, have higher degrees of freedom because they involve a larger number of parameters – with modification indices being inflated as a result. As such, refining the measure until there are no high modification indices would not be reasonable the current study. Secondly, the current scale development process is based on a novel theory that is yet to be thoroughly empirically tested. As such, a decision was made to retain the remaining items to avoid losing theoretically relevant and potentially unique items at such an early stage of development of both the measure and the theory. Nevertheless, future testing and refinement of the scale is recommended as more theoretical evidence proliferates. As noted by Hair and colleagues (2019), there are no unambiguous model fit thresholds, and the values provided in the literature are only intended to be used as guidelines and applied flexibly based on model complexity and theoretical relevance.

The final, post-CFA version of the OAQ refined based the factor analyses consisted of 26 items ($N_{LFC} = 6$, $N_{LRO} = 6$, $N_{LSC} = 7$, $N_{PIE} = 7$). Ordinal alpha and omega values showed each subscale of the OAQ to have satisfactory internal consistency, and the higher-order omega supported the reliability of the OAQ as a holistic tool for assessing maladaptive overcontrol. The reasonable values of model fit indices and the

internal consistency evidence provided preliminary support for the conceptual framework, and thereby Lynch's (2018) theorising regarding the four core deficits of maladaptive overcontrol. The LSC was indicated to have the strongest relationship to the higher order factor, followed by the PIE domain. This suggests that the absence of close, meaningful social relationships, constrained social signalling, and inhibition of emotions and empathic responses lie at the core of maladaptive overcontrol. In turn, the LFC domain had the weakest connection to the higher-order factor. This could suggest that the behaviours and traits within the scope of the LFC deficit contribute the least to the issues of overcontrol. However, considering that the LFC domain has shown the lowest internal consistency of the four domains, as well as the lowest inter-factor correlation with the other three domains in the global EFA, it is also possible that the measure does not accurately reflect the maladaptive behavioural manifestations of overcontrol pertaining to this domain. Considering these results, the researcher decided to compare the face value of the items included within the LFC domain with these of other domains. Looking at the contents of the four subscales, it seems plausible that the items within LFC pool are worded in a less 'maladaptive' way than those in the other three domains. This could mean that the LFC subscale is reflective of self-control tendencies that can be both adaptive and maladaptive, depending on the context in which they occur. Therefore, while the global scale is likely to provide information about maladaptive overcontrolled tendencies, caution is recommended when interpreting standalone LFC subscale scores.

In the EFA study, the inter-factor correlations ranged from weak to strong, providing first indications that the relationships between the domains of maladaptive overcontrol may be more nuanced than what the scale items account for. Still, the satisfactory fit of the higher-order model indicated by the results of the CFA study indicated that relationships between the factors representing the four deficits of overcontrol can be, at least to a certain extent, meaningfully explained by the higher order factor of maladaptive overcontrol. Interestingly, however, the modification indices revealed within the final CFA model indicated a level of covariance between some the four domains that could not be accounted for by the higher order factor, indicating imperfections within the theoretical framework. Because Lynch's (2018) theoretical model is novel, it is possible that more intricate relationships between the deficits and overcontrol exist that are not yet empirically understood or accounted for

in the theoretical model. This result provides a potential explanation for the lower-than-expected correlations between factors indicated by the EFA results. These findings highlight the need for further research that would shed more light on the nature of the relationships between the four core deficits and potential corrections that may need to be applied to the theoretical model, and, subsequently, the OAQ. Unfortunately, considering potential issues associated with overly complex models and insufficient statistical power, specifying additional paths between the domains to test the relationships in more detail was not viable at this time.

6.7.2 Strengths and limitations.

The recruitment for the set of studies was limited to participants residing in English-speaking countries that fall within the WEIRD framework – the UK, the USA, Ireland, Canada, Australia, and New Zealand. In previous studies, recruitment was open to fluent English speakers worldwide to allow for cross-cultural efficacy of the instrument. Still, a vast majority of participants who took part in the studies, while from diverse backgrounds, resided in the UK or the USA. Moreover, the research underpinning the measure development process was conducted predominantly in the listed countries. It was necessary to recognise that the new scale inevitably carried the influence of the social, cultural, and economic characteristics of these countries. As such, a decision was made to limit recruitment accordingly, in hopes of achieving both linguistic consistency and a balance between sensitivity, specificity, and methodological rigour necessary to create a valid and reliable instrument. Nevertheless, future linguistic adaptation of the instrument and cross-cultural validation studies are strongly encouraged.

The EFA and CFA studies were conducted on separate samples of 600 participants. This was to prevent overfitting and include generalisability of the results, in line with the recommendations in literature (e.g., Boateng et al., 2018; Fokkema & Greiff, 2017; Hair et al., 2019). However, due to the time-limited nature of the current programme of research, the data for both the EFA and the CFA study were collected at the same time. This means that following item rejections in the EFA study, participant responses to the rejected items needed to be removed, potentially introducing response bias. Nevertheless, the results of the CFA largely aligned with the results of the global EFA, with two out of three fit indices meeting criteria for the

acceptable fit from the beginning, and few items having to be rejected to further improve model fit. As such, the cross-validation between the EFA and the CFA can be considered a strength of the design. Further, both EFA and CFA were conducted using model estimation and factor rotation methods appropriate for use with ordinal data. Similarly, internal consistency was assessed using methods appropriate for use with ordinal data and in a manner appropriate for multidimensional scales, strengthening the design of the studies.

The data for both studies was collected using a 91-item version ($N_{LFC} = 16$, $N_{LRO} = 23$, $N_{LSC} = 27$, $N_{PIE} = 25$) of the OAQ. Inclusion of lengthy measures and lengthy set of measures in research is often unavoidable (Bowling et al, 2022). However, literature suggests that such lengthy questionnaires have the potential to also introduce response bias and result in lower data quality, lower response rates, and lower completion rates (Bowling et al, 2022; Rolstad et al., 2011). The completion and response rate were not an issue in the current study. This is likely because of the data being collected on Prolific – a platform dedicated to data collection that allows for participants to voluntarily sign up and complete studies in exchange for monetary incentives. The financial gain is larger for longer studies, as the participants are paid pro rata. Such incentives are likely to additionally motivate participants, resulting in high response and completion rates (Abdelazeem et al., 2023). Moreover, in the current study, as per Prolific policy, the accuracy of participant responses was ensured with attention check questions. Data from participants who did not successfully complete at least two out of three attention check questions were removed, hopefully ensuring higher quality data.

6.7.3 Conclusions.

The results of the factor analyses provided support for the conceptual framework developed based on Lynch's (2018) Neurobiosocial Theory for Disorders of Overcontrol, thereby providing some preliminary evidence for the validity of the theoretical model. The domain specific EFAs allowed for the selection of the most appropriate items to assess each of the deficits. The global EFA modelling allowed to ensure that any items that have significant cross-loadings between the four domains or load on an additional factor unaccounted for in the model are rejected. The CFA study allowed to test the theoretically driven higher-order model and to cross-validate the

EFA findings. The new 26-item version of the OAQ has shown acceptable model fit, acceptable to good internal consistency of each of the subscales, and good composite reliability.

7. Locating maladaptive overcontrol within the wider literature and further psychometric evaluation the OAQ.

7.1 Introduction

The two studies discussed in the previous chapter used factor analyses to provide preliminary evidence of construct validity of the OAQ, as well as provided information about the internal consistency reliability of the scale. The current study aims to explore the relationships between the latent construct of maladaptive overcontrol measured by the OAQ and other relevant psychological constructs, as well as test the test-retest validity of the OAQ. The paragraphs that follow first discuss the functions of and controversies around exploring relationships between different psychological constructs with reference to validity. Further analysis of the psychometric properties of the scale is also discussed. Lastly, the chapter describes the aims and hypotheses for the current study, with theoretical rationale for each of the predicted relationships.

7.1.1 Relationships between constructs and the scale's validity.

An important step in the scale development and validation process is establishing how the latent construct, in the way that it is operationalised, fits with other constructs – i.e., how it relates to other psychological constructs that are posited to be theoretically linked¹¹. Findings regarding these relationships are usually referred to as validity evidence, due to the assumption is that a valid measure of a construct should allow researchers to, theoretically, predict its' relationships to other constructs (Hughes, 2018). These relationships are most spoken about with reference to convergent and discriminant or divergent validity – despite little consensus on the meaning of these words. Convergent validity has most often been defined as a type of construct validity concerned with the degree to which participants' scale scores correlate with their scores on measures of conceptually similar constructs, or conceptually related constructs (Colton & Covert, 2007; Hughes, 2018; Murphy & Davidshofer, 2005). The term divergent validity, often used interchangeably with

¹¹ Of note, the researcher acknowledges that criterion validity evidence should also be obtained where possible. This involves comparing a newly developed scale to a different, 'gold standard' scale designed to measure the same construct – if one is available (Colton & Covert, 2007). However, Lynch's (2018) theoretical framework that underpinned the development of the OAQ is novel, and a gold standard measure of maladaptive overcontrol that could be used for this purpose does not exist. As such, criterion validity is not considered further within this thesis.

discriminant validity, has often been understood as a type of construct validity aiming to evidence that two constructs are conceptually distinguishable (or even completely unrelated), or a type of validity aiming to evidence that the scale only measures the construct of interest and not any other construct (Rönkkö et al., 2020). This has been a source of confusion for many aspiring psychometricians.

Interestingly, however, a compelling argument that has been put forward is that these concepts, however defined, have very little to do with *validity*. Validity, by definition, is an umbrella term concerned with whether the scale measures what it intends to measure appropriately and meaningfully (Borsboom et al., 2004; Hughes, 2018; Kelly, 1927) – and while establishing how different constructs relate to each other is crucial to theory building and advancing psychological knowledge, simply observing positive or negative correlations or the lack of correlations between variables does not, per se, indicate whether a scale appropriately and meaningfully measures the desired construct (Borsboom et al., 2004; Hughes, 2018). In the words of Hughes (2018, p. 758):

“[...] completing an intelligence test should require the use of intelligence, and thus differences in test responses between persons would be the result of differences in intelligence. [...] Observing a positive correlation between a measure labeled intelligence and educational achievement actually tells you nothing about what is actually measured.”

There were several criticisms of this view. For example, researchers argued that a scale cannot be validated independently of the interpretation of scale scores and its intended purpose in mind (Hughes, 2018; see e.g., Hubley & Zumbo, 2011; Kane, 2009; Sireci, 2007). As such, because the understanding of relationships between constructs can help ensure that the test is interpretable and fit for purpose, it therefore constitutes part of validity evidence. Consequently, there is certainly no consensus as to the meaning of validity in relation to relationships between constructs that scales measure, and the debate ought to continue.

Nevertheless, in light of this open debate, it is necessary for the researcher to take a stance on the issue. The current thesis adopts the view that relationships between

scale scores and scores on measures should serve primarily to facilitate theory building – especially in the case of novel theories that have not yet been extensively researched, such as the Neurobiosocial Theory for Disorders of Overcontrol. Novel theoretical frameworks are bound to have some inconsistencies and flaws – and these can only be appropriately addressed if extensive research is conducted regarding the construct of interest and the constructs that appear theoretically related. Extending the evidence base lies at the very heart of theory development and is only possible when valid and reliable measures of constructs are available. However, the initial versions of measures can only be grounded in the initial versions of the theory – and as such, it is likely that both the theory and the measure will require refinement as new evidence emerges. Thus, findings that go against the initial hypotheses should not immediately be treated as evidence of poor validity of the scale and lack of credibility of the theory, but rather serve as an indication that future investigation is needed. This iterative process is fundamental to scientific advancement.

Further, many accepted views may be erroneous or poorly defined – with the discussion on quadratic versus linear view of self-control being one of the examples. The lack of a comprehensive framework that would account for the idea of excessive self-control, alongside a lack of an accurate and meaningful assessment measure, has led to the idea of maladaptive overcontrol being scientifically overlooked despite numerous studies over the years offering proof of construct. Additionally, when a single label is used to describe constructs with different operational definitions, as it is in the case of self-control and related constructs, there is a risk that different measures claiming to assess the same construct may actually be tapping into different underlying constructs, oftentimes leading to contradictory results. As such, it is not just novel theories that may be flawed and require revisions, but also more established ones. As such, discrediting a theory, a construct, or a scale and labelling based on unexpected findings does not appear to be a sensible approach. Rather, it should lead to further investigation of the topic, as new and interesting insights may emerge.

The current study focuses on obtaining preliminary evidence on the relationships between maladaptive overcontrol and other constructs that are posited to be conceptually linked to maladaptive overcontrol. However, it will do so to provide the first building blocks for researching and developing the understanding of

maladaptive overcontrol by locating the construct within the wider literature rather than for the purpose of obtaining additional validity evidence. Additionally, the inter-factor correlations within the OAQ are also considered, as means of cross-validating the results from previous, factor-analytic studies.

7.1.2 Maladaptive overcontrol and the links to other constructs.

The OAQ, ego-control, and ego-resilience.

As part of the current study, the scale developer aimed to compare Lynch's (2018) theorising, as measured by the OAQ, with another framework that endorses a quadratic view of self-control and recognises that overcontrol can be problematic – i.e., the theorising on ego-resilience and ego-control. Block's (1971) and Block and Block's (1980) conceptualisation of self-control is the closest to Lynch's. Still, there are several key differences between Lynch's Neurobiosocial Theory for Disorders of Overcontrol and Block and Block's theorising on ego-control and ego-resilience.

The literature on ego-control and ego-resilience has a strong focus on developmental influences on self-control and gives little attention to the role of biological predispositions. Further, the empirical studies within this framework are predominantly exploratory and concentrate on understanding correlations between different personality types that differ on the dimensions of ego-control and ego-resilience and other psychological traits. In contrast, Lynch's model provides a comprehensive account on how maladaptive overcontrol may be developed and maintained over time based on decades of previous research, information of hypothesised mechanisms that function to reinforce maladaptive overcontrolled coping, and integrates nature, nurture, and coping elements within the framework. This provides basis not only for typological and correlational studies, but also for investigations into the mechanisms driving maladaptive overcontrol. Moreover, Block and Block's description of ego-overcontrol does not capture the possible behavioural manifestations quite as comprehensively as Lynch's theory – for instance, it does not mention threat and reward sensitivity. As such, Lynch's maladaptive overcontrol and Block's ego-overcontrol cannot simply be considered synonymous.

Nevertheless, there is certain overlap in the descriptions of Lynch's (2018) maladaptive overcontrol and Block's (1971) ego-overcontrol. Both frameworks

propose that highly overcontrolled individuals are excessively constrained, rigid, inhibited, and guarded, and that they are characterised by high perseverance, a tendency to diligently plan and avoid of ambiguity, detail-focused processing of information, and a strong preference for delaying gratification. Further, both Lynch (2018) and Block and Block (1980) posited that adaptive functioning requires the ability to alter one's level of self-control to changing circumstances - referred to as flexible control by the former and as ego-resilience by the latter. Both frameworks also suggest that highly undercontrolled and highly overcontrolled individuals will be lacking in such ability – a claim recently supported by the results of a study by Isaksson, Ghaderi, Wolf-Arehult, and Ramklint (2021).

In accordance with the above argument, in the current study, it is hypothesised that maladaptive overcontrol – as described by Lynch (2018) and measured by the OAQ – will demonstrate convergence with ego-overcontrol and highlight maladaptive overcontrol and ego-undercontrol as distinct constructs. Further, it is predicted that individuals high in maladaptive overcontrol will present with low levels of ego-resilience, supporting the notion that they experience difficulties appropriately modifying their level of self-control to changing circumstances.

The OAQ and behavioural inhibition and activation.

There is considerable overlap between Lynch's (2018) theorising on maladaptive overcontrol and the way behavioural and trait manifestations of behavioural inhibition and activation systems are conceptualised within the reinforcement sensitivity theory (RST; see e.g., Gray, 1991; Gray & McNaughton, 2000). The RST is a well-known and widely researched physiological model of personality (Carver & White, 1996). The two neurobiological systems originally distinguished by the RST – known as the behavioural inhibition and the behavioural activation systems – have been posited to reflect one's sensitivity to positive (reward) and negative (punishment) reinforcers, motivate approach, avoidance, and withdrawal behaviours, and influence emotion regulation (Carver, 2004; Kimbrel et al., 2007; Quilty & Oakman, 2004; Serrano-Ibáñez et al., 2018).

The behavioural activation system (BAS), also known as the approach system, has been conceptualised as one's sensitivity to potential incentives and activated in the

presence of potentially rewarding stimuli (Carver, 2004; Fowles, 1993). Previous studies have linked higher BAS sensitivity with higher positive affect and extraversion, but also higher impulsivity, anger-out, aggression and frustration, and externalising psychopathology, including bipolar disorder and substance use disorders (Alloy et al., 2009; Carver, 2004; Heubeck et al., 1998; Johnson et al., 2017; Leone & Russo, 2009; Loxton & Dawe, 2001; Quilty & Oakman, 2004; Smits & Kuppens, 2005). The behavioural inhibition system (BIS), on the other hand, has been conceptualised as a system sensitive to punishment, threat, and novelty (Carver & White, 1996). It is also known as the withdrawal system, as it has been shown to be triggered in the presence of stimuli perceived as threatening, and result in inhibitory, avoidant, and withdrawal behaviours and heightened anxiety (Carver, 2004; Davidson, 1992, 1998; Fowles, 1993; Gray, 1994). Higher BIS sensitivity has been empirically linked to higher negative affect and neuroticism, anger-in, and internalising psychopathology, including lifetime diagnoses of anxiety and depression (Heubeck et al., 1998; Johnson et al., 2017; Kasch et al., 2002; Smits & Kuppens, 2005; Sportel et al., 2013).

The sensitivity of the BIS and BAS vary between individuals, and these individual differences are most commonly psychometrically tested using the Behavioural Activation and Inhibition (BAI) measures designed by Carver and White (1996). The BAI is comprised two scales – the Behavioural Activation Scale (BAI-BAS) and the Behavioural Inhibition Scale (BAI-BIS) – designed to capture individual differences in the sensitivity of the behavioural inhibition and activation systems, as conceptualised within the original RST model. The paragraphs that follow consider these scales in relation to Lynch's (2018) conceptualisation of maladaptive overcontrol.

The BAI-BIS is a unidimensional scale which accounts for individual differences in sensitivity of the behavioural inhibition system (Carver & White, 1996). It reflects the differences in reactivity to punishment, threat, and novelty. Lynch (2018) theorised that maladaptively overcontrolled individuals would present with a high threat sensitivity and high behavioural and emotional inhibition, as well as be more likely exhibit internalising behaviour problems, including treatment-resistant anxiety. He also hypothesised that they would avoid novelty and uncertainty, as well as be socially withdrawn. These characteristics directly map onto how a highly

sensitive behavioural inhibition system is conceptualised within the RST, as described in the above paragraph. As such, in the current study, it was hypothesised that higher extent of issues associated with maladaptive overcontrol, as measured by the OAQ, would be linked to higher sensitivity of the behavioural inhibition system, as measured by the BAI-BIS.

The BAI-BAS, in turn, is comprised three components pertaining to (1) sensitivity to reward cues (Reward Responsiveness), (2) seeking out novel and exciting experiences (Fun Seeking), and (3) willingness to determinedly pursue goals (Drive). In relation to reward responsiveness, in his framework, Lynch (2018) hypothesised that highly overcontrolled individuals would present with diminished reward sensitivity – particularly in social contexts. As such, it was also hypothesised that higher levels of maladaptive overcontrol would be linked to lower reward responsiveness on the BAI-BAS. Further, Lynch posited that maladaptively overcontrolled individuals would be avoidant of spontaneity and novelty, which is associated with their posited preference for structure, order, and predictability, and high threat sensitivity. Due to this, it was also expected that highly overcontrolled individuals would be less likely to seek out new, exciting experiences and activities for enjoyment and relaxation purposes, as seen through the lenses of the BAI-BAS. Lastly, in relation to goal pursuit, Lynch (2018) posited that highly overcontrolled people can be expected to persist at tasks in pursuit of their goals due to high perfectionism and the wish to be seen in a positive light by other people. However, the items on the BAI-BAS Drive subscale do not appear to appropriately capture the idea of persistent, goal-directed effort over long periods of time in a manner relevant to Lynch's (2018) theorising. These items appear to instead capture the tendency to relentlessly pursue one's goal, but not necessarily in a way that would be expected of highly overcontrolled, perfectionistic, inhibited individuals. They reflect relentless pursuit of desires, somewhat impulsive and relentless, with little regard for the well-being of others (e.g., 'If I see a chance to get something I want, I move on it right away.' or 'When I go after something I use a "no holds barred" approach.'). Maladaptively overcontrolled individuals, due to their hypothesised high impulse control, strong sense of social obligation, a tendency towards overtolerating distress, and preference for delayed gratification, may be unlikely to endorse these items. However, because the items do not explicitly specify whether they refer to

sacrifice of self or others, or to immediate or distant goals, a decision was made to include the subscale as an exploratory element.

Of note, the researcher acknowledges that an alternative model – the revised-RST (Gray & McNaughton, 2000) – has been proposed, which has distinguished the fight-flight-freeze fear (FFFS-Fear) system from the BIS system, and conceptualised BIS to act as a conflict mediator between the BAS and the FFFS-Fear (see e.g., Heym et al., 2008; McNaughton & Corr, 2008; Poythress et al., 2008). Nevertheless, the original BAI measures developed by Carver and White (1994) that are in accordance with the original BIS/BAS model are still in wide use. While a wider discussion on the RST versus revised-RST is beyond the scope of this thesis, the original BIS item pool appears to be an appropriate to measure both threat and inhibitory sensitivity of maladaptively overcontrolled individuals as theorised by Lynch (2018). As such, a decision was made to use the original BAI scales.

The OAQ and the dimensions of impulsive behaviour.

Impulsivity is a construct closely linked to self-control, with impulsive behaviour believed to involve inhibitory control failure (McHugh & Balaratnasingam, 2018). Previous literature indicated that self-control, conceptualised in line with the linear view, was negatively related to impulsivity – leading researchers to view impulsivity as the opposite of high self-control, and effectively almost equating it with maladaptive undercontrol (Duckworth & Kern, 2010; Mao et al., 2018). However, given that across studies, the correlations between self-control and impulsivity varied in magnitude, others argued that while the constructs are closely related, impulsivity is not synonymous with low self-control (Mao et al., 2018; see Ludwig et al., 2013).

Impulsivity has also been extensively studied in relation to other, both adaptive and maladaptive, personality traits and the resulting behaviours – however, conflicting findings were often reported (see e.g., Mao et al., 2018; McHugh & Balaratnasingam, 2018; Reynolds et al., 2006; Cyders et al., 2014). The mixed results have been attributed to the issues around conceptualising and, consequently, measuring impulsivity. Namely, various conceptualisations of impulsive behaviour and scales corresponding to these have been put forward over the years, and the lack of consensus

as to the best way of defining and measuring the construct resulted in inconsistent findings (Cyders et al., 2014).

Attempting to solve this issue, a new, multidimensional model of impulsive behaviour was proposed, developed based on factor-analytic studies involving various measures of impulsivity (Cyders et al., 2014). It initially included four facets: urgency, (lack of) perseverance, (lack of) premeditation, and sensation seeking (Whiteside & Lynam, 2001). Later, however, two distinct types of urgency – positive urgency and negative urgency – have been recognised, with the model now known as the UPPS-P (Cyders & Smith, 2007). The model has since become very prominent among researchers studying impulsive behaviour in the context of personality, with the UPPS-P measure of impulsive behaviour (Lynam et al., 2006) and its shorter version, the S-UPPS-P (Cyders et al., 2014), becoming perhaps the most widely used psychometric measures of impulsivity. The UPPS-P short- and long-form scales have often been used alongside self-control and impulse control measures, and the UPPS-P facets have been suggested to account for different types of risky behaviours and psychopathological outcomes, many of which were also closely related to self-control (see e.g., Cyders et al., 2007; Fischer et al., 2008; Gray et al., 2019; Johnson et al., 2017; Savvidou et al., 2017; Tran et al., 2018).

Considering the recent developments around overcontrol, the relationship between self-control and impulsivity ought to be examined also in the context of the quadratic view of self-control. Consequently, this is one of the aims of the current study. A decision was made to employ the UPPS-P model of impulsive behaviour, due to its popularity and relevance in the context of personality pathology and mental health disorders (see Um et al., 2018). In line with the Neurobiosocial Theory of Disorders of Overcontrol, it is theorised that maladaptively overcontrolled individuals are high in inhibitory, initiatory (Hamilton, 2021), behavioural, and emotional control, low in reward sensitivity, perseverant, devoted to careful planning, and unlikely to act spontaneously, without thinking, or in a rash manner (Lynch, 2018). As such, it was hypothesised that impulsive behaviour in relation to all five facets within the UPPS-P model will be negatively related to maladaptive overcontrol.

The OAQ and linear models of self-control.

As part of establishing the construct validity of the OAQ, it is crucial to attempt to situate the construct of maladaptive overcontrol also within the broader self-control literature that endorses a linear view of the construct. This can help extend the understanding of how maladaptive overcontrol relates to undercontrol and to desirable, optimal self-control.

As discussed in Chapter 2, the strength model of self-control (Baumeister et al., 1994) and the General Theory of Crime (Gottfredson & Hirschi, 1980) assert that low self-control underlies an array of adverse life outcomes, including complex psychopathology and criminal behaviour. Both theories imply that high self-control equates to good self-control and reject the idea of problematic overcontrol. As such, the BSCS (Tangney et al., 2004) and the LSCS (Grasmick et al., 1993) – the two most used self-control scales guided by the respective frameworks – claim to assess a spectrum from insufficient self-control (or undercontrol) to *high* self-control. Conversely, the current thesis argues that these scales assess a spectrum from undercontrol to *optimal* self-control and lack items that would describe behaviours specific to maladaptive overcontrol that are, at the same time, distinct from undercontrol.

Nevertheless, the scores indicating optimal self-control on these scales can be expected to present with a level of convergence to the OAQ. This is because highly overcontrolled people are likely to endorse items that describe typically associated with adaptive self-control – e.g., report the ability to easily resist impulses, delay gratification, prioritise work over leisure when necessary, control their emotions in public, or take pride in their iron self-discipline. Given that these scales endorse a linear view of self-control and lack items assessing maladaptive overcontrol behaviours but include items signifying behaviours associated with optimal self-control that overcontrolled individuals may endorse, it is reasonable to expect a positive relationship between optimal self-control on these scales and high overcontrol on the OAQ. Further, of interest here is also the relationship between low self-control on the BSCS and LSCS and maladaptive overcontrol on the OAQ, and whether the two linear scales indeed appropriately measure low self-control. An adverse relationship between the two would demonstrate the distinctiveness of undercontrol from overcontrol when

measured by the BSCS and LSCS, and thereby evidence that the scales can indeed accurately assess maladaptive undercontrol. In the previous studies, the two linear scales were used to assess a spectrum from low to high self-control. In accordance with this, the current study hypothesised that convergence would be found between optimal self-control on the BSCS and the LSCS and maladaptive overcontrol on the OAQ due to the overlap in some of the traits. Consequently, negative correlations would be found between maladaptive overcontrol measured by the OAQ and maladaptive undercontrol as measured by both the BSCS, and the LSCS and all six of its components, distinguishing between the two constructs.

7.1.3 Test-retest reliability.

The overarching aim of the current programme of research aims to develop a psychometric measure that would reflect a personality construct of maladaptive overcontrol. Personality traits are defined as personality traits as patterns of thoughts, feelings, and behaviours (Allport, 1961) and can be adaptive or maladaptive. They remain relatively stable over lifespan but can slowly evolve in response to environmental stimuli and psychological interventions (e.g., Bleidorn et al., 2022; Roberts et al., 2017). Due to their relative stability, they have been shown to be able to predict various life outcomes (see e.g., Roberts et al., 2007). As such, good personality measures reflect this stability of traits and allow researchers to predict future outcomes based on the scores. Consequently, to ensure that the OAQ measures stable patterns of personality, it is necessary to evidence that the OAQ scale scores do, in fact, remain relatively stable over time.

The type of reliability that evidences the stability of scores over time is the test-retest reliability (Murphy & Davidshofer, 2005). It is crucial to evidencing appropriate psychometric properties of a newly developed scale, in that it allows to ensure that the scale measures a trait-like, stable construct and is not overly sensitive to state fluctuations (McCrae et al., 2011). This type of reliability is examined through administering the scale of interest to the same sample of participants on (at least) two separate occasions and comparing the scores using correlational analysis. Higher correlation coefficients indicate a better stability of scores over the given period. There is no ‘gold-standard’ interval period between administrations, and the applied interval period between administrations across literature vary from just a few days to several

years in longitudinal studies (Polit, 2014; Quadri et al., 2013). A few weeks has been indicated as optimal to ensure stability without allowing the time for major personality change to occur (De Vries et al., 2016).

7.1.4 Aims and hypotheses.

The aim of the current study was two-fold:

- (1) to situate maladaptive overcontrol within the wider literature pertaining to self-control and constructs related to self-control, through exploring its relationships with these constructs,
- (2) to further evaluate the psychometric properties of the OAQ, looking specifically at inter-factor correlations and test-retest reliability.

It was hypothesised that maladaptive overcontrol, as conceptualised by Lynch (2018), would present a level of convergence with ego-overcontrol (Block, 1971; Block & Block, 1980) and behavioural inhibition (Carver & White, 1996). It was also expected that maladaptive overcontrol would be negatively related to low self-control as conceptualised within two most popular linear models of self-control – the strength model (e.g., Baumeister et al., 1994; 1998) and the General Theory of Crime (Gottfredson & Hirschi, 1980), as well as with the impulsive behaviour on the five dimensions distinguished within the S-UPPS-P (Lynam, 2013), and the Reward Responsivity and Fun Seeking domains of behavioural activation (Carver & White, 1996). The relationship between the drive domain of behavioural activation was also explored, however, no specific hypotheses as to the direction of the relationship have been made.

Lastly, psychometric properties of the OAQ were further examined. Inter-factor correlations within the OAQ were re-examined as means of cross-validating the results from the factor-analytic studies. Four-week test-retest reliability of the OAQ was also tested. A 4-week interval was applied to minimise the likelihood of participants remembering their previous responses, while remaining short enough so that a permanent change in personality is unlikely to occur. It was also practical considering the time constraints of this PhD programme. It was expected that, as a personality measure, the OAQ would present with stability of scores over time.

7.2 Methods

The study plan was submitted for review by the Schools of Business, Law and Social Sciences Research Ethics Committee (reference ID: 1536443). The application included documents such as participant information sheet, consent forms, and debrief forms – all of which were developed in accordance with the British Psychological Society guidelines and internal Nottingham Trent University guidelines. A favourable ethical opinion was granted for the conduct of the study.

6.2.1 Design.

A two-part quantitative study was conducted:

Part 1. The OAQ was presented to participants alongside a battery of scales designed to assess potentially related constructs to examine the relationships between maladaptive overcontrol and these constructs, as well as inter-factor correlations within the OAQ.

Part 2. The OAQ was presented to a subsample of participants who completed Part 1 of the study after a 4-week follow-up period to examine the test-retest reliability of the questionnaire.

Participants

Data were collected via Prolific. Prolific policies regarding participant payment and the amount of funding secured for participant incentives were considered when deciding on the sample size alongside sample sizes previously used in similar studies. The aim was to collect 300 responses to the validation part of the study, and approximately 200 responses to the test-retest part of the study from the validation study participant pool.

Maladaptive overcontrol in relation to other psychological constructs.

Part 1. To take part in the study, the participants needed to be 18 years old or over and fluent in English, and resided in English-speaking countries that fall within the Western, Educated, Industrialised, Rich, and Democratic (WEIRD) framework – the UK, the US, Ireland, Canada, Australia, and New Zealand. The balanced sample option was chosen when setting up the study on Prolific, allowing for collection of

approximately 50% female and 50% male responses. For a response to be considered valid and the participant approved, the participants must have consented to take part in the study, indicated that they were residing one of the countries of interest, that they spoke fluent English, successfully submitted the survey, and failed no more than one out of the four attention check questions (e.g., *'This is an attention check, please choose disagree slightly.'*).

As planned, total of 300 valid responses were collected in this part of the study. Each participant who submitted a valid response received an incentive for participation in line with Prolific policy. The sample included 149 females (49.67%) and 148 males (49.33%), with three people not disclosing their sex (1.00%). The average age of participants was 37.37 years old ($SD = 13.81$, range: 18-86 years old). One hundred and eleven participants resided in the UK (37.00%), 70 in Canada (23.33%), 65 in Australia (21.67%), 29 in New Zealand (9.67%), 19 in the USA (6.33%), and six in Ireland (2.00%). All participants declared fluency in English, with it being the first language for 248 participants (82.67%) and a preferred language for 46 participants (15.33%). For more information on the demographic characteristics sample, please refer to Appendix I.

Part 2. To take part in the second part of the study, participants must have submitted a valid response to the first part of the study. Again, the balanced sample option was chosen when setting up the study on Prolific, allowing for collection of approximately 50% female and 50% male responses. For a response to be considered valid and the participant approved, the participants must have consented to take part in the study, indicated that they were residing one of the countries of interest, that they spoke fluent English, successfully submitted the survey, and failed no more than one out of the two attention check questions (e.g., *'This is an attention check, please choose 'Agree.'*).

A total of 184 valid responses were collected in this part of the study. Each participant who submitted a valid response was rewarded for their time in line with Prolific policy (£1.00). The sample included 92 females (50.00%) and 90 males (48.91%), with two people not disclosing their sex (1.09%). The average age of the participants was 40.22 years old ($SD = 13.99$, range: 20-86 years old). Seventy-one participants resided in the UK (38.58%), 44 in Australia (23.91%), 38 in Canada

(20.65%), 19 in New Zealand (10.33%), seven in the USA (3.80%), and five in Ireland (2.72%). All participants declared fluency in English, with it being the first language for 155 participants (84.24%) and a preferred language for 26 participants (14.13%). For more information on the demographic characteristics sample, please refer to Appendix J.

7.2.3 Measures.

Overcontrol Assessment Questionnaire (OAQ)

A 26-item, four-factor version of the OAQ refined based on factor analyses was included in the study. The scale includes six numerical response options ranging from *Strongly disagree* (1) to *Strongly agree* (6), and an additional *No opinion* option intended to be coded as missing data. There is one reverse-scored item. Scores on the OAQ can range between 26 and 156, with higher scores indicating higher levels of traits associated with maladaptive overcontrol, as theorised by Lynch (2018). Importantly, low scores are not designed to indicate undercontrol. It was the only psychometric scale to be completed twice, both in the validation and the test-retest part of the study.

Ego Undercontrol-13 Scale (EUC-13; Isaksson, Ghaderi, Wolf-Arehult, & Ramklint, 2021)

The original, 37-item version of the scale, designed to measure the spectrum from under- to overcontrol on a single dimension, can be traced back to Block and Block (1980). It was first evaluated in published research by Letzring and colleagues (2005). The researchers evaluated the psychometric properties of the scale and did not find them to be satisfactory (Letzring et al. 2005). As such, the measure, despite being the only one to assess under- and overcontrolled tendencies, has rarely been used in research. Recently, however, Isaksson, Ghaderi, Wolf-Arehult, and Ramklint (2021) revisited the scale and attempted to validate it for use in Swedish general population. The psychometric properties were, again, not found to be satisfactory. To improve these, the researchers developed a short, 13-item version – the EUC-13. The measure assesses three factors that aim to measure different behaviours conceptualised to differentiate between under- and overcontrol: (1) *Uninhibited behavior* (5 items, e.g., ‘*When I get bored, I like to stir up some excitement.*’), (2) *Planful and conscientious*

behavior (5 items, e.g., reverse-scored ‘*I like to stop and think things over before I do them.*’) and (3) *Socially restrained behavior* (3 items, e.g., reverse-scored ‘*I find it hard to make small talk when I meet new people.*’). Items are scored on a scale from 1 (*Disagree very strongly*) to 4 (*Agree very strongly*), with lower mean scores indicating higher overcontrolled tendencies and higher mean scores indicating higher undercontrolled tendencies. The internal consistency of the global scale ($\alpha = .71$) and the former two subscales ($\alpha = .76$ for both) have been found to be satisfactory. The internal consistency of the *Socially restrained behavior* subscale, however, was less than satisfactory ($\alpha = .51$), and as such, the sole use of the subscale was advised against. The results also indicated satisfactory test-retest validity ($r = .86$ for the global scale and .75-.84 across subscales). Given the lack of an English version of the questionnaire with satisfactory psychometric properties, a decision was made to use the EUC-13 despite its development in a non-English speaking population, in the current study. This is due to shorter length and better psychometric properties when compared to the original English version. Permission for use was obtained from Dr Martina Isaksson.

Ego Resilience Scale (ER89; Block & Kremen, 1996)

The unidimensional, 14-item ER89 scale was developed to measure the construct of ego resilience proposed by Block and Block (1980) and defined as the ability to adapt one’s level of control depending on circumstances. Items are scored on a scale from 1 (*Disagree very strongly*) to 4 (*Agree very strongly*), with higher mean scores indicating representing higher adaptability (e.g., ‘*I quickly get over and recover from being startled.*’). The scale has been shown to have satisfactory internal consistency ($\alpha = .76$) and 5-year test-retest reliability of .67 for females and .51 for males (adjusted for attenuation). The scale was accessed via APA PsycTESTS database, which indicated that no permission was necessary to use the scale for non-commercial research purposes.

S-UPPS-P Impulsive Behavior Scale (Lynam, 2013)

The 20-item S-UPPS-P is a short version of a 59-item UPPS-P developed by Lynam and colleagues (2006) to assess impulsive behaviour in the context of the Big Five personality traits (Costa & McCrae, 1992). The S-UPPS-P consists of five subscales: (1) *Sensation Seeking* (the tendency to engage in exciting and novel activities; e.g., reverse-scored ‘*I quite enjoy taking risks.*’), (2) *Negative Urgency* (the tendency to act

impulsively when negative affect is present; e.g., reverse-scored ‘*When I am upset I often act without thinking.*’), (3) *Lack of Premeditation* (the tendency to act on an impulse, without prior planning and without considering the potential consequences of the behaviour; e.g., ‘*I usually think carefully before doing anything.*’), (4) *Lack of Perseverance* (the tendency to quickly give up on tasks when they become tiring or disinteresting; e.g., ‘*I finish what I start.*’), and (5) *Positive Urgency* (the tendency to act impulsively when positive affect is present; e.g., ‘*I tend to lose control when I am in a great mood.*’). Each subscale contains four items, all of which are scored on a scale 1 (*Agree strongly*) to 4 (*Disagree strongly*). Higher total subscale score indicates a higher level of impulsive behaviours assessed by the subscale. Considering the significant length of the battery of measures, a decision was made to use the time-efficient short-form version in the current study. The S-UPPS-P offers satisfactory psychometric properties (subscale $\alpha = 0.74\text{--}0.88$), similar factor structure, inter-correlations, and little loss of shared variance when compared to the long form UPPS-P, as well as a much shorter completion time (Cyders et al., 2014). While test-retest reliability of the short version was not examined by Cyders and colleagues (and the author is unaware of any studies to date that have examined this specifically), translated versions of the questionnaire been shown to have satisfactory test-retest reliability (e.g., test-retest correlations of .84-.92 in the French version; Billieux et al., 2012). Permission for use was obtained from Dr Melissa Cyders.

Behavioral Activation and Inhibition Scales (BAI; Carver & White, 1996)

The BAI scales have been developed to measure trait sensitivity level of the behavioural activation (Behavioral Activation Scale, or BAS) and inhibition (Behavioral Inhibition Scale, or BIS) systems, in line with the RST (Gray, 1982; Gray, 1991). The BAS contains 13 items across three dimensions measuring different aspects conceptualised to be reflect activatory sensitivity: (1) *Reward Responsiveness* (reflecting one’s sensitivity to reward cues; 5 items, e.g., ‘*When I get something I want, I feel excited and energized.*’), (2) *Fun Seeking* (reflecting one’s willingness to seek out and pursue novel and exciting activities and experiences; 4 items, e.g., ‘*I crave excitement and new sensations.*’), and (3) *Drive* (reflecting one’s willingness to determinedly pursue their desired goals; 4 items, e.g., ‘*I go out of my way to get things I want.*’). The 7-item BIS is a unidimensional measure, with the items reflecting one’s tendency for anxiety, fear, and worry (e.g., reverse-scored ‘*Even if something bad is*

about to happen to me, I rarely experience fear or nervousness.'), as well as sensitivity to criticism (e.g., *'Criticism or scolding hurts me quite a bit.'*). Both the BAS and the BIS are scored on a scale from 1 (*Strongly disagree*) to 4 (*Strongly agree*), with higher total scores reflecting a higher sensitivity of the corresponding system. Carver and White (1996) have reported acceptable internal consistencies for BIS (Cronbach's $\alpha = .74$), Reward Responsiveness (Cronbach's $\alpha = .73$) and Drive (Cronbach's $\alpha = .76$), and an alpha value slightly below the cut-off for Fun Seeking (Cronbach's $\alpha = .66$). They have also found test-retest correlations of .66 for BIS, .66 for Drive, .59 for Reward Responsivity, and .69 for Fun Seeking. The scales were accessed via APA PsycTESTS database, which indicated that no permission was necessary to use these for non-commercial research purposes.

Low Self-Control Scale (LSCS; Grasmick, 1993)

The LSCS is a 24-item¹² measure developed to assess six components (Impulsivity, Simple Tasks, Risk Seeking, Physical Activities, Self-Centered, and Temper; 4 items per component) of a single personality trait of low self-control in the context of criminology, and in line with Gottfredson and Hirchi's (1990) General Theory of Crime. The items are rated on a scale from 1 (*Strongly disagree*) to 4 (*Strongly agree*), with higher total global score indicating lower self-control. The original paper reported a Cronbach's alpha value of .81 for the scale, indicating good internal consistency. Test-retest reliability analyses were not performed at the time. As discussed in Chapter 2, later studies reported mixed results regarding dimensionality and psychometric properties of the scale, repeatedly put the scale's asserted cultural universality into question, and prompted some researchers to argue that different components of the scale should also be considered separately and advise against the use of the global score (e.g., Pechorro et al., 2023). Nevertheless, the scale is by far the most widely used scale for measuring self-control in the context of criminology and forensic behaviour, and even referred to as the gold standard (Pechorro et al., 2023). The scale was accessed

¹² Some studies use a 23-item version of the scale. The inconsistency stems from Grasmick et al. (1993) removing one of the items (*'I seem to have more energy and a greater need for activity than most other people my age.'*) from the original, 24-item scale due to little contribution to the validity and reliability of the scale (Pechorro et al., 2023). However, the 24-item version was available on the APA PsycTESTS database and as such, utilised in the current study.

via APA PsycTESTS database, which indicated that no permission was necessary to use the scale for non-commercial research purposes.

Brief Self-Control Scale (BSCS; Tangney et al., 2004)

The BSCS is a shorter, 13-item version of a 36-item Self-Control Scale, designed by Tangney and colleagues (2004) to measure individual differences in trait self-control. It is one of the most used self-control scales (Manapat et al., 2021), aligned with the strength model of self-control (Baumeister et al., 1994), as well as Carver and Scheier's (1981, 1982, 1998) conceptualisation of self-regulation feedback loops. While studies support both unidimensional and multidimensional structures of the scale, a global score across all the items is usually considered (Lindner et al., 2015; Manapat et al., 2021). Higher total scores on the scale have been understood to indicate higher trait self-control, with lower total scores indicating lower trait self-control. The scale was found to have satisfactory psychometric properties – Tangney and colleagues (2004) reported high internal consistency of the BSCS (Cronbach's $\alpha = .83-.85$) and test-retest reliability of .87. Permission for use was obtained from Dr June Tangney.

Demographics.

A demographic questionnaire was also included in the study, with questions regarding the following variables: age, country of origin, country of residence, ethnicity, sex assigned at birth, gender identity, marital status, education level. It also included a question on whether English is a participant's first or preferred language. It was included in both the validation and the test-retest study to account for the possibility that participants' circumstances have changed between the two times the OAQ was administered.

7.2.4 Procedure.

The survey was advertised exclusively on Prolific and filled in by participants using Qualtrics. Participants were first presented with a *captcha* to screen for AI-assisted entries. Upon successful completion of the scan, participants were presented with an information sheet, a consent form, and a demographic screener. If a participant did not consent to taking part in the study, indicated residing in a country other than the UK, USA, Ireland, Canada, Australia, or New Zealand, and/or indicated that they did not speak fluent English, they were informed that they do not meet the inclusion criteria

and redirected back to Prolific. On Qualtrics, after consenting to take part and filling in the demographic information, participants were presented with the battery of measures in the first part of the study, and the OAQ only in the test-retest part of the study. For all questionnaires, the order of items was randomised to avoid order effects. At the end of the study, participants were presented with a Prolific completion code necessary to confirm study participation and collect the incentive, as well as a download link for the debrief file. All participants who submitted valid responses were included in the analyses.

7.3 Data collation and analysis

Microsoft Excel and RStudio were used for data collation and cleaning. Statistical analyses were conducted using RStudio in version 2023.06.0+421 running under macOS Ventura 13.0.1. Packages used included base (v4.2.2), dplyr (v1.1.0), Hmisc (v5.1-1), misty (v0.6.2), naniar (v1.1.0), psyntur (v0.1.0), stats (v4.2.2), and tidyverse (v2.0.0).

7.3.1 Maladaptive overcontrol in relation to other constructs.

The data file was downloaded from Prolific, and unnecessary columns were removed. The qualitative response labels were recoded as numerical values, in line with each questionnaire's scoring instructions. On the OAQ, all 'No opinion' responses were recoded as missing values. Where a questionnaire included reverse-scored items, the scoring was also adjusted.

There were 88 missing datapoints in the questionnaire data (excluding demographics), accounting for .52% of all values. Eighty-three per cent of all missing values were accounted for by the OAQ responses only, with a total of 73 missing values (.93% of all responses to the OAQ) and a total of 42 participants with at least one missing value (14.00% of all participants). This was due to the No opinion response being treated as missing data, which accounted for 72 (98.63%) of all missing values in the OAQ. Little's MCAR test (Little, 1988) was conducted to inspect the missing data patterns. Thirty-nine patterns were found, and the results indicated no significant differences between the patterns ($\chi^2(4862) = 4915.53, p = .292$), and thereby provided no evidence against the hypothesis that the data are MCAR.

Total or mean scores for each measure and subscale were subsequently calculated, depending on original scoring instructions. Considering the very low percentage of missing data and the results of the MCAR test, a decision was made to use the weighted sum scores to account for missing data when calculating the total scores. Descriptive statistics were calculated. Inspection of skewness values (ranging from -.73 to .53) and kurtosis values (ranging from 2.35 to 3.32) indicated a non-normal distribution of the data (see Appendix K). This was followed by obtaining a correlation matrix using Spearman's (1910) rank correlation coefficients and Holm's (1979) *p*-value correction method. Classification of effect sizes cited by Ellis (2010) was endorsed, with correlation coefficient values of < .10 representing a negligible correlation, values of .10-.29 representing a weak correlation, values of .30-.49 representing a moderate correlation, and values > .50 representing a strong correlation. *P*-values < .05 indicated significant results and rejection of the null hypothesis. The ordinal alpha values (Zumbo et al., 2007) were also calculated for the scales and subscales to examine internal consistency, with values of .70 and above indicating satisfactory reliability.

7.3.2 Test-retest reliability.

The follow-up data file was downloaded from Prolific, and unnecessary columns were removed. The qualitative response labels were recoded as numerical values in line with the OAQ's scoring instructions, and the 'No opinion' responses were recoded as missing values. Where necessary, items were reverse scored.

There were 37 missing datapoints in the OAQ data, all of which were a result of recoding the 'No opinion' option as a missing value. The missing values accounted for 5.47% of all values. Little's MCAR test (Little, 1988) was conducted to inspect the missing data patterns. Twenty patterns were found, and the results indicated no significant differences between the patterns ($\chi^2(463) = 470.16, p = .399$), and thereby provided no evidence against the hypothesis that the data are MCAR.

The Prolific IDs of participants who completed the follow-up questionnaire were matched with their responses with the initial responses to the OAQ from the validation dataset. Subsequently, total scores were calculated for each timepoint (*T1* – initial completion; *T2* – repeated completion after four weeks). Here again, the

weighted sum scores to account for missing data when calculating the total scores. Descriptive statistics were also calculated. Inspection of skewness values (ranging from -.76 to .04) and kurtosis values (ranging from 2.20 to 3.91) indicated a non-normal, mesokurtic distribution of the data. Spearman's (1910) rank correlation coefficients and *p*-values corrected using Holm's (1979) method were used to examine the test-retest reliability of the OAQ and the subscales. Scale and subscale ordinal alpha values (Zumbo et al., 2007) were also calculated to examine the internal consistency once again.

7.4 Results

7.4.1 Maladaptive overcontrol and other psychological constructs.

Descriptive statistics.

Descriptive statistics presenting central tendencies, variability, and range of total or mean scores of the participant sample on scales and scale subscales included in the battery of measures are presented in Table 7.1.

Table 7.1: Descriptive statistics for the scale and subscale scores of the participants who took part in the first part of the study.

Scale	Subscales	M	SD	Md	Min.	Max.
OAQ	Global	102.70	16.88	105.00	42.64	140.00
	LFC	27.24	4.47	27.00	11.00	36.00
	LRO	24.21	5.78	25.00	7.00	36.00
	LSC	22.63	6.85	23.00	7.00	40.25
	PIE	28.59	6.53	29.58	9.00	42.00
EUC-13*	Global	2.72	0.45	2.77	1.46	3.77
	Uninhibited behavior	2.28	0.59	2.20	1.00	3.80
	Planful conscientious behavior	3.02	0.70	3.20	1.00	4.00
	Socially restricted behavior	2.96	0.78	3.00	1.00	4.00
ER*	-	2.84	0.41	2.86	1.57	4.00
S-UPPS-P	Negative Urgency	8.57	2.68	9.00	4.00	15.00
	Lack of Perseverance	7.76	2.27	8.00	4.00	14.00
	Lack of Premeditation	7.12	2.20	7.00	4.00	14.00
	Sensation Seeking	9.00	2.24	9.00	4.00	16.00
	Positive Urgency	8.05	2.38	8.00	4.00	16.00
BAI-BIS	-	21.45	3.86	21.00	11.00	28.00
BAI-BAS	Reward Responsiveness	15.83	2.16	15.50	9.00	20.00
	Drive	10.27	2.54	10.00	4.00	16.00
	Fun Seeking	10.36	2.16	10.00	4.00	16.00
LSCM	Global	51.68	8.71	52.00	28.00	77.00
	Impulsivity	8.57	2.38	8.00	4.00	16.00
	Simple Tasks	9.50	2.52	9.50	4.00	16.00
	Risk Seeking	8.04	2.74	8.00	4.00	16.00
	Physical Activities	9.03	2.63	9.00	4.00	16.00
	Self-Centered	7.99	2.25	8.00	4.00	15.00

	Temper	8.55	2.67	9.00	4.00	16.00
BSCS	-	40.36	8.56	40.00	20.00	65.00

Note. *Mean rather than total scores were calculated for scales marked with an asterisk.

Internal consistency of the OAQ and inter-factor correlations.

The results indicated ordinal alpha values of .89 for the global OAQ scale, .80 for the LFC subscale, .79 for the LRO subscale, .86 for the LSC subscale, and .85 for the PIE subscale. Positive correlations were found across the OAQ, presented in Table 7.2. The results indicated strong correlations between the global OAQ score and all subscale scores. Correlations between subscale scores ranged from weak to strong, with the weakest correlations between the LFC domain and the LSC and the PIE domains, and the strongest correlation between the LSC domain and the PIE domain. All correlations were significant at the .05 level.

Table 7.2: Inter-factor correlations within the OAQ.

	OAQ Global		OAQ-LFC		OAQ-LRO		OAQ-LSC	
	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>
OAQ-LFC	.50	<.001						
OAQ-LRO	.68	<.001	.33	<.001				
OAQ-LSC	.80	.005	.16	<.001	.39	<.001		
OAQ-PIE	.73	.005	.16	<.001	.25	<.001	.53	<.001

EUC-13.

The results indicated an ordinal alpha value of .62 for the EUC-13 scale, with subscale values of .75 for the Uninhibited Behavior, .44 for the Planned Conscientious Behavior, and .31 for the Socially Restricted Behavior. Values for all but the Uninhibited Behavior subscale were below satisfactory, indicating poor internal consistency of the EUC-13.

The correlations between the OAQ and the EUC-13 and their subscales are presented in Table 7.3. Several weak to moderate negative correlations were found, supporting the hypothesised direction of the relationship between the scales, and providing partial evidence for conceptual similarity of maladaptive overcontrol as defined by Lynch (2018) and ego-overcontrol as defined by Block and Block (1980). Significant negative correlations were found between:

- (1) the global OAQ scores and the global EUC-13 scores (weak correlation), as well as the EUC-13 Socially Restricted Behavior (moderate correlation),
- (2) the OAQ-LFC and the global EUC-13 scores (weak correlation), as well as the EUC-13 Planned Conscientious Behavior (moderate correlation),
- (3) the OAQ-LRO and the global EUC-13 scores (weak correlation), as well as the EUC-13 Socially Restricted Behavior (moderate correlation),
- (4) the OAQ-LSC and the EUC-13 Socially Restricted Behavior (moderate correlation),
- (5) the OAQ-PIE and the EUC-13 Socially Restricted Behavior (weak correlation).

The remaining correlations were not found to be significant. Due to the poor internal consistency of the EUC-13, the below results should be interpreted with caution.

Table 7.3: Correlations between the OAQ and the EUC-13 scales.

	OAQ Glob.		OAQ-LFC		OAQ-LRO		OAQ-LSC		OAQ-PIE	
	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>
EUC-13 Global	-.14	.017	-.23	<.001	-.15	.008	-.02	.759	-.08	.177
EUC-13 Uninh. Beh.	.11	.067	-.01	.883	-.02	.703	.15	.010	.10	.084
EUC-13 Planned Consc. Beh.	-.08	.191	-.31	<.001	-.01	.805	.07	.220	.20	-.214
EUC-13 Socially Restr. Beh.	-.34	<.001	-.09	.123	-.34	<.001	-.30	<.001	-.20	<.001

Note. Significant results are presented in **bold**.

ER89.

Statistical analysis indicated an ordinal alpha value of .81 for the ER89 scale, evidencing satisfactory internal consistency. The results of the correlation matrix indicated a significant moderate negative correlation between the ER89 scores and the global OAQ scores ($rho = -.44, p < .001$), supporting the direction of the hypothesised relationship between maladaptive overcontrol and ego resilience. Inverse relationship between the ER89 scores and the OAQ scores was also supported at subscale level for the OAQ-LRO ($rho = -.50$ [strong correlation], $p < .001$), the OAQ-LSC ($rho = -.47$ [moderate correlation], $p < .001$), and the OAQ-PIE ($rho = -.20$ [weak correlation], p

< .001) domains. The correlation between the ER scores and the OAQ-LFC subscale scores was not found to be significant ($\rho = -.04, p = .483$).

BAI-BIS.

Statistical analysis indicated an ordinal alpha value of .88 for the BIS scale of the BAI, evidencing satisfactory internal consistency. Convergence between the global OAQ scores and the BIS scores was supported by the results, with a significant moderate positive correlation found ($\rho = .30, p < .001$). Convergence between the BIS and the OAQ was also found at subscale level for the OAQ-LFC scores ($\rho = .19$ [weak correlation], $p = .001$), OAQ-LRO scores ($\rho = .65$ [strong correlation], $p < .001$), and the OAQ-LSC scores ($\rho = .12$ [weak correlation], $p = .039$). However, no significant correlation was found between the BAI scores and the OAQ-PIE scores ($\rho = -.02, p = .690$).

BAI-BAS.

Statistical analysis revealed ordinal alpha values of .80 for the BAS Reward Responsiveness subscale, .89 for the Drive subscale, .79 for the Fun Seeking subscale, indicating satisfactory internal consistency.

The correlations between the OAQ and the BAS subscales are presented in Table 7.4. The results indicated a mix of positive and negative associations, ranging from weak to moderate. In line with the predictions, significant negative correlations were found between:

- (1) the global OAQ scores and the BAS Fun Seeking (weak correlation),
- (2) the OAQ-LRO and Fun Seeking (weak correlation),
- (3) the OAQ-LSC and each of the BAS subscales (weak correlations),
- (4) the OAQ-PIE and each of the BAS subscales (weak correlations).

Contrary to the predictions, a significant positive correlation was found between the OAQ-LFC and the BAS Reward Responsiveness (weak correlation). No hypotheses were previously made regarding the OAQ and the BAS Drive. The exploratory analysis indicated that the OAQ-LRO was significantly negatively correlated the BAS Drive (weak correlation), while the OAQ-LFC was significantly positively correlated with BAS Drive (moderate correlation).

Table 7.4: Correlations between the OAQ and the BAI-BAS subscales.

	OAQ Glob.		OAQ-LFC		OAQ-LRO		OAQ-LSC		OAQ-PIE	
	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>
BAS Reward Responsiveness	-.09	.128	.18	.001	.03	.547	-.21	<.001	-.19	<.001
BAS Drive	-.08	.166	.31	<.001	-.16	.004	-.15	.011	-.12	.036
BAS Fun Seeking	-.15	.008	-.06	.263	-.12	.031	-.14	.018	-.09	.117

Note. Significant results are presented in **bold**.

S-UPPS-P.

Statistical analysis revealed ordinal alpha values of .81 for the S-UPPS-P Negative Urgency subscale, .79 for the Lack of Perseverance subscale, .89 for the Lack of Premeditation subscale, indicating satisfactory internal consistency of these subscales. However, ordinal alpha values of .54 for the Sensation Seeking subscale and .65 for the Positive Urgency subscale were found, indicating internal consistency below the acceptable threshold.

The correlations between the OAQ scale scores and the S-UPPS-P subscale scores are presented in Table 7.5. The results indicated a mix of positive and negative associations, ranging from weak to moderate. In line with the hypotheses, significant negative correlations were found between:

- (1) the OAQ-LFC and the S-UPPS-P Lack of Perseverance (moderate correlation) and Lack of Premeditation (moderate correlation),
- (2) the OAQ-PIE and the S-UPPS-P Lack of Perseverance (weak correlation) and Lack of Premeditation (weak correlation).

Contrary to the predictions, the results indicated convergence between:

- (1) the global OAQ scores and the S-UPPS-P Negative Urgency (weak correlation),
- (2) the OAQ-LRO and the S-UPPS-P Negative Urgency (moderate correlation),
- (3) the OAQ-LSC and the S-UPPS-P Negative Urgency (weak correlation) and Lack of Perseverance (weak correlation).

Table 7.5: Correlations between the OAQ and the S-UPPS-P subscales.

	OAQ Glob.		OAQ-LFC		OAQ-LRO		OAQ-LSC		OAQ-PIE	
	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>
S-UPPS-P Negative Urgency	.22	<.001	.05	.369	.36	<.001	.19	<.001	-.02	.713
S-UPPS-P Lack of Perseverance	-.10	.090	-.44	<.001	.06	.272	.12	.040	-.13	.025
S-UPPS-P Lack of Premeditation	-.09	.101	-.34	<.001	.02	.682	.08	.166	-.12	.044
S-UPPS-P Sensation Seeking	-.00	.987	.01	.905	-.00	.934	.01	.821	-.00	.990
S-UPPS-P Positive Urgency	.00	.974	-.02	.762	-.06	.762	.06	.341	.00	.342

Note. Significant results are presented in **bold**.

The remaining correlations were not found to be significant. Due to the less-than-satisfactory internal consistency, results pertaining to the S-UPPS-P Sensation Seeking and Positive Urgency subscales should be interpreted with caution.

BSCS.

Statistical analysis indicated an ordinal alpha value of .89 for the BSCS scale, evidencing satisfactory internal consistency. The results of the correlation matrix indicated a significant weak negative correlation between the global OAQ scores and the BSCS scores ($rho = -.22, p < .001$), contrary to what was predicted. Similarly, at subscale level, a significant moderate negative correlation was found between the BSCS scores and the OAQ-LRO scores ($rho = -.36, p < .001$), and a significant weak negative correlation was found between the BSCS scores and the OAQ-LSC scores ($rho = -.28, p < .001$). In contrast, a significant weak positive correlation was found between the BSCS scores and the OAQ-LFC scores ($rho = -.13, p < .027$), which was in the direction hypothesised. No significant association was found between the BSCS scores and the OAQ-PIE scores ($rho = -.04, p = .192$).

LSCS.

Statistical analysis indicated ordinal alpha values of .83 for the LSCS scale, with subscale values of .75 for Impulsivity, .79 for Simple Tasks, .85 for Risk Seeking, .79 for Physical Activities, .68 for Self Centered, and .76 for Temper. The results indicated a satisfactory internal consistency for global scale and all but the Self Centered subscales (albeit also approaching the threshold).

The correlations between the OAQ and the LSCS, including and their subscales, are presented in Table 7.6. The results indicated a mix of positive and negative associations, ranging from weak to moderate. In line with the predictions, significant negative correlations were found between:

- (1) the global OAQ scores and the LSCS Physical Activities (weak correlation) and Temper (weak correlation),
- (2) the OAQ-LFC and the LSCS Impulsivity (weak correlation),
- (3) the OAQ-LRO and the LSCS Risk Seeking (weak correlation) and Physical Activities (weak correlation),
- (4) the OAQ-LSC and the LSCS Physical Activities (weak correlation).

Contrary to the hypotheses, the results indicated convergence between:

- (1) the global OAQ scores and the global LSCS scores (weak correlation), the LSCS Simple Tasks (weak correlation) and Self Centered (moderate correlation),
- (2) the OAQ-LFC scores and the LSCS Self Centered (weak correlation) and Temper (weak correlation),
- (3) the OAQ-LRO and the global LSCS scores (weak correlation), as well as the LSCS Impulsivity (weak correlation), Simple Tasks (moderate correlation), and Temper (moderate correlation),
- (4) the OAQ-LSC and the global LSCS scores (weak correlation), as well as the LSCS Impulsivity (weak correlation), Simple Tasks (weak correlation), Self Centered (moderate correlation), and Temper (weak correlation),
- (5) the OAQ-PIE and the LSCS Self Centered (weak correlation).

The remaining correlations were not found to be significant. Due to the less-than-satisfactory internal consistency, results pertaining to the LSCS Self Centered subscale should be interpreted with caution.

Table 7.6: Correlations between the OAQ and the LSCS scale and its subscales.

	OAQ Glob.		OAQ-LFC		OAQ-LRO		OAQ-LSC		OAQ-PIE	
	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>	<i>rho</i>	<i>p</i>
LSCS Global	.21	<.001	.07	.222	.17	.003	.26	<.001	.06	.270
LSCS Impulsivity	.10	.064	-.15	.008	.12	.035	.19	.001	.06	.270
LSCS Simple Tasks	.24	<.001	.02	.762	.40	<.001	.25	<.001	.00	.993

LSCS Risk Seeking	-.07	.242	-.06	.313	-.14	.017	-.02	.780	.01	.911
LSCS Physical Act.	-.14	.012	.06	.295	-.18	.002	-.13	.027	.11	.057
LSCS Self Centered	.33	<.001	.15	.010	.05	.431	.43	<.001	.22	<.001
LSCS Temper	-.22	<.001	.20	<.001	.33	<.001	.25	<.001	.08	.193

Note. Significant results are presented in **bold**.

Other considerations.

The remaining correlation coefficients from the correlation matrix are presented in Appendix L for informative purposes. Due to the unexpected correlation patterns between the OAQ and the LSCS and the OAQ and the BSCS, a decision was made to inspect the scales' associations with the ER89, with psychological inflexibility being central to Lynch's (2018) theorising regarding maladaptive overcontrol. The goal was to examine whether the relationship to ego resilience differed depending on how each scale conceptualises (good) self-control. No significant correlation was found between the LSCM scores and the ER89 scores ($\rho = -.02, p = .761$), while a weak positive correlation was found between the BSCS scores and the ER89 scores ($\rho = -.27, p < .001$). For a comparison, a significant moderate negative correlation reported between the OAQ and the ER89 was previously reported ($\rho = -.44, p < .001$).

7.4.2 Test-retest reliability.

Descriptive statistics presenting central tendencies, variability, and range of total scores on the OAQ and its subscales for participants who took part in both parts of the study are presented in Table 7.7. Strong correlations were found between participants scores at T_1 and T_2 for the global scores ($\rho = .87, p < .001$), the LFC subscale ($\rho = .79, p < .001$), the LRO subscale ($\rho = .82, p < .001$), the LSC subscale ($\rho = .87, p < .001$), and the PIE subscale ($\rho = .83, p < .001$). The results indicated good to excellent test-retest reliability of the OAQ and its subscales.

Table 7.7: Descriptive statistics for the OAQ global and subscale scores of the participants who participated in both the validation and the test-retest reliability parts of the study.

	M		SD		Md		Min.		Max.	
	<i>T1</i>	<i>T2</i>	<i>T1</i>	<i>T2</i>	<i>T1</i>	<i>T2</i>	<i>T1</i>	<i>T2</i>	<i>T1</i>	<i>T2</i>
OAQ Global	101.90	102.19	17.33	17.33	104.00	104.00	42.64	48.00	140.00	139.00
OAQ-LFC	27.35	27.22	4.61	4.43	28.00	27.00	11.00	14.00	36.00	36.00
OAQ-LRO	23.76	23.69	5.92	5.80	24.00	24.00	8.00	7.00	36.00	36.00
OAQ-LSC	22.37	22.67	6.83	7.02	22.00	23.00	8.00	8.00	40.25	38.00
OAQ-PIE	28.38	28.50	6.63	6.53	29.00	29.00	10.00	9.00	42.00	41.00

7.5 Discussion.

The aim of the current study was to explore the relationships between maladaptive overcontrol and other, potentially related psychological constructs, as well as establish the test-retest reliability of the newly developed OAQ. To achieve these objectives, a two-part quantitative study. The OAQ was first administered to a sample of participants alongside a battery of other self-report scales to test the relationship between overcontrol and other relevant psychological constructs. After four weeks, the OAQ was administered again to participants from the same sample, so that stability of scores over time can be evaluated. The results provided partial support for the study's hypotheses, and thereby the Neurobiosocial Theory for Disorders of Overcontrol (Lynch, 2018). The current section offers an interpretation of the study's findings. It considers the inter-factor correlations within the OAQ and its test-retest reliability, as well as the OAQ's relationships with the measures included in the battery. Future directions for research are also proposed.

7.5.1 The OAQ: Internal consistency, test-retest reliability, and inter-factor correlations.

The results of the study indicated satisfactory internal consistency for the OAQ and all its domains, adding credibility to the results of the previous study. Further, good to excellent 4-week test-retest reliability of the OAQ and its subscales was found. The strong agreement between participant scores on the first and the second administration indicates that the scores remain stable over time. This finding suggests that the OAQ indeed measures a trait rather than a state and provide preliminary evidence that the scale can be used for the purpose of repeated evaluations. However, future studies should evaluate test-retest reliability of the OAQ over longer periods of time to ensure its utility in instances where lengthier intervals between administrations are required, such as measuring treatment change.

The correlation matrix indicated strong positive correlations between the OAQ and all its domains, indicating that all the subscales provide meaningful information about elements of the overarching construct of maladaptive overcontrol that they were designed to measure. The weakest correlations were found between the LFC domain and the LSC and PIE domains, suggesting that while there is a meaningful relationship

between these deficits, the degree of association is not as pronounced as that between the other factors. A moderate correlation was observed between the LFC domain and the LRO domain. The strongest correlation was found between the LSC domain and the PIE domain, indicating a high level of interrelation between the two deficits. This pattern of correlations resembles one from the EFA study, adding credibility to the results. Further, the current study's results align somewhat with those of the CFA study, where modification indices highlighted covariance between the LFC and LRO domains, as well as between the LSC and PIE domains, which were unaccounted for by the higher-order construct of maladaptive overcontrol. This suggests the existence of more nuanced interrelationships between the deficits. Specifically, the LFC and LRO domains on the one hand, and the LSC and PIE domains on the other hand, more closely related to bi-directionally, than to the remaining deficits. This cannot be fully explained by the current versions of the theory and the conceptual model. As such, further research on this is needed, which may lead to updates to the theoretical model and to the OAQ.

7.5.2 Maladaptive overcontrol and ego-control.

The study hypothesised that maladaptive overcontrol, as well as each of its core deficits, would be negatively correlated to EUC-13 scores – both indicating convergence between maladaptive overcontrol and ego-overcontrol, and highlighting the differences between maladaptive overcontrol and ego-undercontrol. The results partially supported the hypotheses, and the significant results are summarised below. However, the internal consistency of the EUC-13 was found to be below satisfactory in the current sample, with only the Uninhibited Behavior subscale meeting the desired threshold. This effectively undermines the legitimacy of the results. Consequently, minimal interpretation of the findings was undertaken, and the results should be approached with considerable caution.

In line with the hypotheses, a significant negative correlation was found between the global scores on the OAQ and the global scores on the EUC, likely reflecting the similarities in conceptualisations of problematic overcontrol provided by Block (1971), Block and Block (1980), and Lynch (2018). At a global level, the EUC-13 was also significantly negatively correlated with the LFC and the LRO domains of the OAQ. In turn, the global maladaptive overcontrol scores were negatively correlated

with the EUC-13 Socially Restricted Behavior subscale scores. The Socially Restricted Behavior subscale of the EUC-13 was also, as predicted, significantly negatively correlated with the LRO, LSC, and PIE subscales of the OAQ. Also, in accordance with the study's hypotheses, the Planned Conscientious Behavior subscale of the EUC-13 was significantly negatively correlated only with the LRO subscale of the OAQ, which is likely a reflection of the previously described conceptual similarities between these elements of the two theoretical frameworks (i.e., the preference for planning, structure, order, and certainty). No other significant correlations were found.

Interestingly, no significant correlations were found between the OAQ and the EUC-13 Uninhibited Behavior subscale – the only EUC-13 subscale that presented with satisfactory internal consistency. To attempt to explain this unexpected result, the EUC-13 items pertaining to the subscale were inspected. The researcher noticed that one item – *'At times, I am tempted to do or say something that others would think inappropriate.'* – pertained to the internal thoughts of the respondent rather than their actual behaviour. Considering Lynch's (2018) conceptualisation of maladaptive overcontrol, it would not be surprising to find a highly overcontrolled individual report that they are *tempted* to say inappropriate things – but they would likely never act on these temptations. Similarly, it would be consistent with Lynch's framework for a highly overcontrolled individual to report that their behaviours are frequently misunderstood by other people – meaning that the item *'My way of doing things can be misunderstood or bother others.'* may be reflective of both under- and overcontrolled individuals. As such, the subscale may not appropriately measure maladaptive overcontrolled behaviours and may not accurately reflect the distinction between under- and overcontrol.

The challenge with obtaining satisfactory internal consistency reliability in the current study, as well as in previous studies (see Isaksson, Ghaderi, Wolf-Arehult, & Ramklint, 2021; Letzring et al., 2005) suggests that Block and Block's (1980) theorising – which served as a foundation for the development of the original EUC and its later versions, including the EUC-13 – may be considerably flawed. It could be that the scale does appropriately differentiate between the constructs of undercontrol and overcontrol due to reductionist operationalisation of the constructs, with the descriptions serving merely as clinical descriptors rather than forming a comprehensive

theory. As such, the EUC and EUC-13 may simply not capture a full spectrum of under- and overcontrolled difficulties, especially compared to the much more comprehensive conceptualisation of maladaptive overcontrol presented by Lynch (2018). Further, as previously mentioned, it is unknown when the original EUC was developed and there exists no record detailing how the item pool was generated and refined (Letzring et al., 2005), rendering it impossible to evaluate the rigour of the scale development process. Consequently, the use of the EUC or the EUC-13 in future studies that wish to compare undercontrolled and overcontrolled presentations in English-speaking WEIRD populations is not recommended.

7.5.3 Maladaptive overcontrol and ego-resilience.

It was hypothesised that maladaptive overcontrol as an overarching construct, as well as all four core deficits of maladaptive overcontrol, would be negatively correlated with ego-resilience, as measured by the ER89. The ER89 demonstrated satisfactory internal consistency in the current sample.

Significant negative correlations were found between ego-resilience and maladaptive overcontrol, as well as the LRO, LSC, and PIE deficits – with the magnitude of the correlations ranging from weak to strong. These results indicated that deficits in receptivity and openness, social connectedness, and emotional expressivity, as well as the overall extent of difficulties associated with maladaptive overcontrol, are linked to decreased ability to flexibly adapt to changing circumstances. This is in accordance with Lynch’s (2018) theorising and the study’s hypotheses.

However, contrary to the hypotheses, ego-resilience and the LFC domain were not found to be significantly correlated. This finding appears to align with the interpretation of results of the factor-analytic studies discussed in the previous chapter, again suggesting that the wording of items in the LFC domain may not behavioural manifestations of overcontrol that are *always* maladaptive. Rather, when considered as a standalone subscale, it may be reflective of self-control behaviours that can be both adaptive and maladaptive, depending on the context in which they occur. The LFC items could, therefore, be endorsed by individuals characterised both by adaptive self-control and maladaptive overcontrol, explaining the lack of a correlation between the domain and the ego-resilience scores. Thus, caution must be exercised when

interpreting standalone LFC scores, especially if the deficits in flexible control are not present alongside pronounced deficits in other domains of maladaptive overcontrol.

7.5.4 Maladaptive overcontrol and behavioural inhibition.

It was hypothesised that individuals high in maladaptive overcontrol and more profound difficulties in relation to all four deficits would present with a more sensitive behavioural inhibition system, as measured by the BAI-BIS scale. The BAI-BIS demonstrated satisfactory internal consistency in the current sample.

The results partially supported the study's hypotheses. Convergence was found between the BAI-BIS scores and the global OAQ scores, as well as the LFC, LRO, and LSC subscale scores. This finding suggests that individuals with higher levels of maladaptive overcontrol present with more sensitive behavioural inhibition system – i.e., experience anxiety more profoundly in response to threat and punishment cues, as well as be more likely to avoid and react fearfully to novelty (see e.g., Blackford et al., 2018; Carver, 2004, 2009; Carver & White, 1996; Lahat et al., 2011) – in accordance with Lynch's (2018) theorising. The results also indicate that the link between maladaptive overcontrol and high behavioural inhibition is especially apparent in relation to the overcontrolled preference for planning, structure, and order, as well as profound deficits in receptivity, openness, and social connectedness.

Contrary to the predictions, no significant correlation was found between the BAI-BIS scores and the OAQ-PIE scores. This unexpected result could be a result of the focus of the BAI-BIS on whether an individual experiences the feelings of worry and anxiety in response to various environmental cues and occurrences (e.g., *'I feel pretty worried or upset when I think or know somebody is angry at me.'* or *'Criticism or scolding hurts me quite a bit.'*), rather than on whether and how they are likely to express these emotions. According to Lynch's (2018) theory, maladaptively overcontrolled individuals can experience the feelings of worry and anxiety as intensely as those who are highly undercontrolled, and for similar reasons. The difference between under- and overcontrolled individuals lies in whether and how they express these emotions – with those overcontrolled likely to pervasively inhibit their expression. The items on the BAI-BIS refer to emotional impact of events, but not the expression of the emotions related to those events, which likely explains the lack of a

significant association. It also further differentiates the construct of behavioural inhibition from the construct of maladaptive overcontrol, with only maladaptive overcontrol additionally accounting for inhibited emotional expression.

7.5.5 Maladaptive overcontrol and behavioural activation.

It was hypothesised that individuals high in maladaptive overcontrol and more profound difficulties in relation to all four deficits would present with a diminished reward sensitivity and a decreased tendency to seek out novel and exciting experiences, as measured by the BAI-BAS Reward Responsiveness and Fun Seeking subscales, respectively. No specific hypotheses were made regarding the potential association between maladaptive overcontrol and the pursuit of one's desires as measured by the Drive subscale of the BAI-BAS. The BAI-BAS demonstrated satisfactory internal consistency in the current sample.

Contrary to the predictions, global scores on the OAQ were not significantly correlated with BAS Reward Responsiveness scores. Individuals' reward sensitivity was, however, significantly negatively correlated with the OAQ-LSC and the OAQ-PIE scores – but no significant correlation was found with the OAQ-LRO scores. Lynch (2018) suggested that the diminished reward sensitivity of highly overcontrolled individuals applies primarily to social and emotional contexts. However, it may not extend to other contexts, such as work and school. This is supported by the current findings and aligns with the posited perfectionistic tendencies of maladaptively overcontrolled individuals and their high motivation to excel in structured, goal-directed environments and be seen as accomplished and successful by others. Unexpectedly, a positive correlation was found between the BAS Reward Responsiveness and the OAQ-LFC, in support of the previous suspicions that the LFC may, as a standalone subscale, measure both adaptive and maladaptive behaviours, depending on the context in which they occur. In this case, the positive correlation may reflect that higher reward responsiveness of individuals characterised by adaptive levels of self-control.

As predicted, maladaptive overcontrol was negatively correlated to Fun Seeking, indicating a decreased inclination towards seeking out novel and exciting experiences. This is in line with Lynch's (2018) positing that maladaptively

overcontrolled individuals avoid novelty and change and have a strong preference for structure and order. Significant negative correlations were also found between BAS Fun Seeking and the LRO, LSC, and PIE deficits of maladaptive overcontrol. This finding indicates that avoidance of novelty, uncertainty, and risks in maladaptively overcontrolled individuals extends not only to everyday activities and events, but also to interpersonal relationships and emotional expression. No significant correlation was found between the BAS Fun Seeking scores and the OAQ-LFC scores, possibly once again indicating the LFC item pool may reflect both adaptive and maladaptive behaviours when treated as a standalone subscale.

No hypotheses were previously made regarding the Drive subscale of the BAS. The difficulty is that the subscale seems to reflect relentless pursuit of one's goals with little regard for the potential consequences to other people rather than, as it is claimed, the idea of persistent, goal-directed efforts. The exploratory analysis of the relationship between maladaptive overcontrol and BAS Drive indicated no significant correlation between the subscale and global OAQ scores. A significant negative correlation was found between the BAS Drive and the OAQ-LRO, indicating that individuals with profound deficits in receptivity and openness are less likely to take risks and ruthlessly and selfishly pursue their desired goals. Relentless goal pursuit with little regard for the consequences exposes one to criticism from others – the very thing that maladaptively overcontrolled people wish to avoid. Positive correlations were found between BAS Drive and the OAQ-LFC, the OAQ-LSC, and the OAQ-PIE. These findings indicate that individuals with a stronger preference for structure, order, and predictability, and more deficits in social connectedness and emotional expressivity are more motivated to pursue their goals in a detached, task-oriented manner. It could be that individuals who are generally more detached from other people and their own feelings and highly driven by structure, order, and predictability find it easier to pursue their desired goals relentlessly and with little concern for the potential socioemotional aftermath.

7.5.6 Maladaptive overcontrol and the dimensions of impulsive behaviour.

The study predicted that individuals scoring high on maladaptive overcontrol would have less inclination towards different types of impulsive behaviour, as measured by the five subscales of the S-UPPS-P. Satisfactory internal consistency was found for the

Negative Urgency, Lack of Perseverance, and the Lack of Premeditation subscales of the S-UPPS-P. However, the internal consistency of the Sensation Seeking and Positive Urgency subscales was found to be less than satisfactory in the current sample.

In line with the predictions, the LFC and the PIE subscales of the OAQ were found to negatively correlate with the Lack of Perseverance and the Lack of Premeditation subscales of the S-UPPS-P. This suggests that individuals who have a strong preference for planning, structure, order, and certainty, as well as have profound difficulties with expressing emotions, are inclined to act with premeditation and are likely to persevere at tasks and see them through to the end. These findings align with Lynch's (2018) theorising. However, no significant correlation was found between the Lack of Premeditation subscale of the S-UPPS-P and the LRO and LSC domains of the OAQ, and the Lack of Perseverance subscale of the S-UPPS-P and the LRO domain of the OAQ. These results did not support the study's hypotheses, and they are challenging to explain given little research available on maladaptive overcontrol. It could be that these latent constructs, as defined by their corresponding subscales, share little conceptual overlap, and this is simply reflected by the lack of a significant relationship found in the study. However, more research is needed to better understand these results – potentially with use of alternative scales measuring the dimensions of S-UPPS-P. Further, contrary to the predictions, individuals characterised by social aloofness and disconnectedness, as measured by the LSC subscale of the OAQ, were found to be less likely to persevere at tasks and follow them through. It could be that highly undercontrolled and highly overcontrolled individuals are likely to report similar difficulties with interpersonal relationships, albeit the underlying cause of these difficulties would differ depending on their self-control tendencies. This could explain why individuals who lack in perseverance were found to report difficulties in social interactions. Further research is encouraged to explore this possibility.

In contrast to the predictions, the S-UPPS-P Negative Urgency subscales was found to be significantly positively correlated to the global OAQ scores, as well as the OAQ-LRO and OAQ-LSC subscale scores. No significant correlations were, in turn, found between the S-UPPS-P Negative Urgency and the OAQ-LFC and the OAQ-PIE. To explain these results, the S-UPPS-P Negative Urgency item pool was closely examined, and it became clear that these pertain to the general tendency to ruminate

on negative emotions, with only half of the items making a reference to rash behaviour in response to negative emotions (i.e., *'When I am upset I often act without thinking.'* and *'When I feel rejected, I will often say things that I later regret'*). Further, Lynch (2018) theorised that after prolonged periods of distress in which a maladaptively overcontrolled individual was inhibiting their intense negative emotions, there may come a moment where the individual loses control and has an unexpected and explosive emotional outburst. As such, these two items could potentially still be endorsed by maladaptively overcontrolled individuals who experience emotional leakage and later regret losing control over their emotions and experience feeling of guilt. Similarly, the item *'Sometimes when I feel bad, I can't seem to stop what I am doing even though it is making me feel worse.'* could be endorsed by maladaptively overcontrolled individuals in reference to their tendency to overtolerate distress.

No significant correlations were found between the OAQ and its subscales and the Sensation Seeking and the Positive Urgency subscales of the S-UPPS-P. This can likely be explained by the poor internal consistency of the two S-UPPS-P subscales. Further studies are needed to better understand the relationship between the psychological constructs of sensation seeking and positive urgency and maladaptive overcontrol, as well as the overall relationship between maladaptive overcontrol and different conceptualisations of impulsivity.

7.5.7 Maladaptive overcontrol and the linear view of self-control.

The OAQ endorses a quadratic view of self-control – it assumes that both insufficient and excessive self-control can be problematic. However, the two scales that are most widely used to measure self-control – the BSCS (Tangney et al., 2004) and the LSCS (Grasmick et al., 1997) endorse a linear view, and argue that self-control is only problematic when it is too low. They interpret problematic self-control in accordance with the strength model of self-control and the General Theory of Crime, respectively. While they claim to measure a spectrum of low to high self-control, they lack items that capture behaviours and traits typical to overcontrol and distinct from undercontrol.

In the current study, it was hypothesised that a level of convergence would be found between higher self-control scores on the BSCS and the LSCS and scores

indicating higher levels of maladaptive overcontrol on the OAQ. Consequently, negative correlations would be found between maladaptive overcontrol and maladaptive undercontrol. This is because even problematically overcontrolled individuals may be likely to endorse items which indicate higher levels of self-control where no distinction between adaptive levels and maladaptively high levels of self-control is made (e.g., *'I am good at resisting temptation.'* and *'People would say that I have iron self-discipline.'* of the BSCS), as well as disagree with items that indicate high undercontrolled tendencies (e.g., *'Excitement and adventure are more important to me than security.'* and *'I don't devote much thought and effort to preparing for the future.'* of the LSCS).

The BSCS.

The BSCS presented with satisfactory internal consistency reliability in the current sample. Positive correlations between the BSCS and the OAQ were expected.

Contrary to what was predicted, the results of the correlation matrix indicated a negative correlation between maladaptive overcontrol at the global level and the BSCS scores. Similarly, at the subscale level, negative correlations were found between the BSCS scores and the OAQ-LRO and the OAQ-LSC scores. An interpretation of these results in accordance with the conceptualisations of the BSCS in previous literature would be that maladaptive overcontrol, as measured by the OAQ, presents with convergence with low self-control, or undercontrol, as measured by the BSCS. An alternative explanation – and less perplexing – explanation is that the BSCS measures neither the spectrum from low self-control to high self-control, nor the spectrum from low self-control to optimal self-control. Rather, it measures a spectrum from maladaptive self-control to adaptive self-control. This is plausible, as some of the BSCS items designed to capture undercontrol could also likely be endorsed by highly overcontrolled individuals – but for different reasons. For example, the item *'I say inappropriate things.'* could be endorsed by problematically overcontrolled due to the feeling that they do not always know what to say or how to behave in social contexts, or the item *'I wish I had more self-discipline.'* could be endorsed due to the wish to have even more perfect self-control.

A positive correlation was found between the BSCS scores and the OAQ-LFC scores. Considering previous indications that the LFC domain may be endorsed by individuals characterised by adaptive levels of self-control, this result aligns with the researcher's interpretation of the results. The lack of a significant association between the BSCS scores and the OAQ-PIE scores, in turn, can likely be explained by the lack of items that specifically relate to emotional experiences and carry emotional significance within the BSCS.

The LSCS.

The LSCS presented with satisfactory internal consistency at both global and subscale level, with the exception of the subscale pertaining to the Self Centered component, internal consistency of which was slightly below the desired threshold. Negative correlations between the OAQ and the LSCS were expected. However, the results indicated a mix of positive and negative correlations.

Globally, maladaptive overcontrol was positively correlated with trait low self-control as defined by the LSCS, which was against the hypothesis. The finding could indicate that the LSCS does not measure undercontrol, but rather captures maladaptive expressions of both low and high self-control. As such, it may be that the experiences undercontrolled and overcontrolled individuals that result from maladaptive levels of self-control are, to some extent, similar, but the causes of these difficulties and deficits may be different. Indeed, upon closer inspection, several items designed to capture high undercontrol appear likely to be endorsed by maladaptively overcontrolled individuals. For example, the Simple Tasks item *'When things get complicated, I tend to quit or withdraw.'* could be endorsed by them due to the tendency to withdraw from uncomfortable social situations and shut down, and the Self Centered item *'I'm not very sympathetic to other people when they are having problems.'* could be endorsed by maladaptively overcontrolled individuals due to their posited low empathic orientation. Considering the potential issues with the LSCS and its conceptualisation of low self-control, even the interpretations of the results that were in line with the study's hypotheses should be taken with caution. Positive correlations against the predictions were also found between Contrary to the hypotheses, the results indicated convergence between the global OAQ and the LSCS Simple Tasks and Self Centered subscales, the OAQ-LFC and the LSCS Self Centered and Temper subscales, the

OAQ-LRO and the global LSCS and the Impulsivity, Simple Tasks, and Temper subscales, the OAQ-LSC and the global LSCS and the Impulsivity, Simple Tasks, Self Centered, and Temper subscales, and the OAQ-PIE and the LSCS Self Centered subscale. However, currently, in the light of limited empirical research in relation to maladaptive overcontrol, a more detailed interpretation of these results is not feasible.

Considering the potential issues with the LSCS and its conceptualisation of low self-control, even the interpretations of the results that were in line with the study's hypotheses should be interpreted with caution. Interestingly, despite an overall positive correlation between the OAQ and the LSCS, the global OAQ scores correlated negatively with the Physical Activities and the Temper components of the LSCS in line with the predictions, indicating that people with higher levels of maladaptive overcontrol prefer mental over physical activities and are less likely to lose their temper when angry or upset. The hypothesis relating to the former finding was guided by Gottfredson and Hirschi's (1980) conceptualisation of low self-control rather than directly by Lynch's (2018) theorising. However, it could be that overcontrolled individuals are purpose-driven in relation to physical activity, and they prefer to plan activities ahead and engage primarily in ones that they perceive as advantageous and efficient rather than spontaneous and frivolous. The latter finding, in turn, provides support for Lynch's positing on the tendency for maladaptively overcontrolled individuals to internalise emotions and inhibit their expression. The LFC domain of the OAQ correlated negatively with Impulsivity, which was in line with the predictions, and indicated that individuals with a preference for structure and order are generally more likely to plan and think ahead rather than act spontaneously in pursuit of immediate indulgence. Negative correlations in line with the predictions were also found between the LRO domain and the Risk Seeking and the Physical Activities dimensions of the LSCS. The former of these findings suggests that the individuals who have more pronounced deficits associated with overcontrolled receptivity and openness are less likely to engage in risky activities just for the fun of it and regardless of the consequences – which directly supports Lynch's positing. The latter indicates that these individuals prefer mental over physical activities, which, again, cannot be straightforwardly related to Lynch's theory. It may be that they simply enjoy mental activities due to them being more predictable and carry less danger compared to physical activities. Lastly, a significant negative correlation in line with the study's

predictions was found between the LSC domain of the OAQ and the Physical Activities component of the LSCS. It may be that individuals characterised by higher levels of overcontrol prefer mental activities as they can easily be performed in isolation and involve minimal social interaction.

The remaining correlations were not found to be significant and did not support the study's hypotheses. Overall, the perplexing mix of positive, negative, and no correlations revealed a complex and nuanced relationship between the constructs that the OAQ and the LSCS measure. Importantly, the results highlighted that both the LSCS and the OAQ measure multidimensional constructs, and that the LSCS conceptualisation of low self-control may need to be updated. Alternatively, it may be that the OAQ does not measure a construct perpendicular to the LSCS. As discussed in Chapter 2, the LSCS has previously been widely criticised, one of the reasons being that it has been designed in line with conceptualisation of self-control endorsed by the General Theory of Crime – i.e., viewed self-control as a unidimensional construct with some identifiable components that are not meaningful on their own (Hamilton, 2021; Piquero et al., 2000; Venables et al., 2018). As such, dimensionality, construct validity, and universality of the scale have all been questioned (Gibson et al., 2010; Higgins, 2007; Marcus, 2004; Pechorro et al., 2023; Vazsonyi & Belliston, 2007; Ward et al., 2018). Moreover, while some of the six dimensions of low self-control proposed by Gottfredson and Hirschi have been widely accepted to be crucial to the construct of low self-control, some have been viewed as less relevant (Pechorro et al., 2023), or argued to be consequences of low self-control rather than components of it (Hoyle & Davisson, 2017). These arguments suggest that the LSCS may not actually measure undercontrol – or, at least, not accurately – and again suggests that the linear view of self-control may be erroneous. The current findings underscore that more research is needed to understand what exactly differentiates overcontrol from undercontrol.

The links to ego-resilience.

Due to the unexpected correlation patterns between the OAQ and the LSCS and the OAQ and the BSCS, the scales' associations with ego-resilience (measured by the ER89) were additionally considered. The goal was to examine whether the relationship to ego resilience differs depending on how each scale conceptualises self-control. No correlation was found between the LSCM scores and the ER89 scores, suggesting that

the construct that the LSCM measures does not have any significant association to psychological flexibility or resilience. This suggests that the LSCM may, indeed, not capture the construct of low self-control as it claims to capture. A weak positive correlation was, however, found between the BSCS scores and the ego resilience, indicating that individuals who score higher on the BSCS are more likely to be able to flexibly adapt to changing circumstances than those scoring lower. This suggests that high scores on the BSCS may indeed capture optimal, adaptive self-control. To reiterate, a negative correlation was found between maladaptive overcontrol and ego-resilience, indicating little flexibility and adaptability of highly overcontrolled individuals. These very different results indicate that each of the three scales measure, at least to some extent, different constructs, or at least different aspects of the same construct of self-control.

7.5.8 Conclusion.

To conclude, the current study's findings supported some of the hypotheses and consequently provide support for certain elements of the Neurobiosocial Theory for Disorders of Overcontrol (Lynch, 2018). Maladaptive overcontrol was linked to lowered ability to adapt the level of self-control to changing circumstances and showed some convergence with the construct of ego-overcontrol (Block, 1971; Block & Block, 1980). Parts of Lynch's (2018) theorising regarding the links between maladaptive overcontrol and characteristics associated with lowered behavioural inhibition and, to some extent, also heightened behavioural activation, were supported. However, complex relationships were revealed between maladaptive overcontrol and the dimensions of impulsive behaviour, as well as other, linear conceptualisations of self-control. The results of the current study strongly suggest that while there exist great conceptualisations of optimal, flexible self-control – but that maladaptive levels of self-control, both high and low, are not yet well understood. The results highlighted maladaptive overcontrol as multifaceted. It clearly underscored the need to move away from the linear view that the more self-control, the better. As such, further empirical investigation across various populations is warranted, and should focus on understanding how maladaptive overcontrol develops, is maintained, and manifests, as well as how relates to wider psychological literature and differs from undercontrol.

8. Conclusions and reflections

Despite the idea of excessive self-control being present in the literature for several decades, a lack of a comprehensive and testable theoretical framework that would explain how maladaptive overcontrol may develop and be maintained was lacking. In 2018, based on a synthesis of decades of translational research, a novel Neurobiosocial Theory for Disorders of Overcontrol was proposed by Thomas Lynch. However, a lack of a quick and reliable measure that would allow for a confident identification of overcontrolled individual has been hindering research progress in the area. The overarching aim of this thesis was to develop a new, theoretically driven self-report questionnaire that would allow for an efficient identification of issues associated with maladaptive overcontrol, as conceptualised within Lynch's theoretical framework. It was envisioned that the scale would allow researchers to identify the presence of problems associated with maladaptive overcontrol, but also the extent of problems associated with each of the four core deficits that Lynch proposed in his model.

This final chapter presents a summary of the methods employed and synthesizes the studies conducted to achieve the research objective and their results. The final version of the OAQ is presented, complete with the instructions for respondents, scoring procedures, and recommendations for scale administration. The sample limitations and the scale's generalisability and utility in its current form are considered, along with proposed future directions for research. Additionally, the chapter shortly reflects on the importance of embracing the iterative approach to scale and theory development and on navigating the psychometric literature as an early career researcher without previous experience in scale development.

8.1 Synthesis of study methods and results.

A robust psychometric approach guided the development and validation of the OAQ. The scale development was theoretically driven, and methodological decisions were guided by an extensive review of the literature pertaining to psychometric measure development.

8.1.1 Conceptual framework.

Chapter 3 described the first task of the scale development process – determining the conceptual framework and the boundaries of the construct that would guide the scale development process. Within the conceptual framework, the four core deficits of maladaptive overcontrol identified by Lynch (2018) were mapped out as domains of the construct, in accordance with the posited multidimensional nature of maladaptive overcontrol. Conceptualising these four deficits as the scale’s dimensions allowed to identify the presence and extent of maladaptive overcontrol globally, but also to understand which of the deficits may have the most pronounced negative impact on each respondent’s everyday functioning. This is beneficial for future research, as it allows to understand the relationships between particular deficits of overcontrol and other psychological constructs and psychopathology in more detail. Further, it is crucial in the context of treatment planning, delivery, and evaluation, as it allows for the treatment goals could to be tailored to the client’s specific difficulties and clinical change evaluated more comprehensively. Literature-informed decisions about the functional properties of the scale were also made at this stage. Based on a review of literature, the decision was made to develop a Likert-type scale with six numerical response options and an additional ‘No opinion’ option that would be treated as a skipped question/missing data.

8.1.2 Item generation.

Chapter 3 also presented the process of item generation. A total of 190 items were initially generated across the four domains through literature review and theoretically-driven focus groups with the potential scale respondents and centred around both the posited observable behavioural manifestations of the deficits and internal states associated with them. Combining the deductive and inductive methods of item generation allowed to ensure that the item pool was as comprehensive as possible (Boateng et al., 2018). This was especially important given that while based on translational research, the assumptions of the novel theory are yet to be thoroughly empirically tested. Incorporating both observable behaviours and internal states aimed to provide a holistic view of maladaptive overcontrol, capture its multifaceted nature (Lynch, 2018), and ensure that the items can provide theoretically meaningful and practically applicable insights regarding the nature of overcontrolled difficulties. The

relatively large number of items within the initial pool increased the likelihood that, following a rigorous refinement process at later stages of the project, enough well-functioning items would remain to ensure a robust and meaningful final version of the scale (Boateng et al., 2018). Overall, this comprehensive approach to item generation increased the likelihood that the final scale would accurately reflect the scope and complexity of the construct of maladaptive overcontrol.

8.1.3 Content validation.

Chapter 4 focused on establishing content validity. The contents of the item pool were validated using the expert judgement method. It was ensured that among the experts were researchers familiar with the theoretical model and clinicians with experience of treating of overcontrolled clients to ensure both theoretical and clinical relevance of the contents (Beck, 2020; Boateng et al., 2020; Fernández-Gómez et al., 2020). All experts were presented with the conceptual framework, including both conceptual and operational definitions of maladaptive overcontrol and each of the four deficits, to ensure the consistency in understanding of the key constructs among the experts (Beck, 2020). The experts were also given the option to be credited in the research outputs as means of recognising their contribution and enhancing accountability and transparency of the content validation process (Beck, 2020).

Quantitative ratings were utilised in the study, and appropriate content validity statistics were calculated to guide the refinement of the item pool (Alamanasreh et al., 2019; Halek et al., 2017; Koller et al., 2017). The criteria on which the items and domains were assessed were carefully defined ensure consistency in the interpretation and application of the ratings (Beck, 2020). There was also space for the experts to provide qualitative comments in case they wished to clarify their ratings or provide specific recommendations for the refinement of the items and the scale, as an attempt to combine the benefits of both quantitative and qualitative approaches to expert judgement (Newman et al., 2017). Any items modified or added to the pool following the initial round of ratings were rated again to validate the changes made.

The methods chosen for the conduct of the study prioritised time-efficiency to minimize expert burden while, by applying a high level of methodological rigour, ensuring that the quality of the study was not compromised. Overall, the robust

approach utilised within the study helped to ensure the scale's theoretical and clinical relevance. Following the refinements based on the data, the item pool was reduced to 94 items across the four domains.

8.1.4 Pre-testing of the scale.

In the subsequent study, the refined version of the item pool was administered to a sample of potential respondents using online cognitive interviews, with the scale further refined based on the interview contents. The cognitive interviews utilised established and rigorous methods to ensure high-quality data, combining both cognitive probes and the think-aloud approach (Fernández-Gómez et al., 2020; Shiyabola et al., 2019; Willis, 2004; Willis & Artino, 2013).

The study helped establish that the respondents share a common understanding of the meaning of the items that is in line with the theoretical framework, warranting comparable and interpretable scores (Boateng et al., 2018; Conrad and Blair, 1996; Drennan, 2003; Reynolds et al., 1993). Further, the study allowed the scale developer to ensure that the items are written in a clear and easily understood language. This is crucial to minimising the respondents' cognitive load and ensuring that the scale is suitable for use with individuals across educational backgrounds and with varying cognitive abilities (Clark & Watson, 2016; Conrad and Blair, 1996; Drennan, 2003; Lezner, 2012; Reynolds et al., 1993). Following refinements based on the content of cognitive interviews, the item pool was reduced to 91 items across the four domains.

8.1.5 Factor analyses.

Two factor-analytic studies were conducted to extract factors, perform item reduction, and establish model fit. Despite having some preconceptions about the structure of the data based on the theoretical model, an exploratory approach was used as a precursor to a confirmatory approach due to the limited previous empirical work around the theory (Gerbing & Hamilton, 1996; Hurley et al., 1997). The data for the studies were collected simultaneously, with the total participant sample split into two to allow for cross-validation of the results and avoid overfitting the model to the data (Boateng et al., 2018; Fokkema & Greiff, 2017; Hair et al., 2019).

The first study utilised EFA to explore the structure of the scale and guide item reduction (Hair et al., 2019). Factor and item retention were driven both by the statistical analyses and the theoretical background (Hair et al., 2019). The EFA study allowed for retention of only the most meaningful items in the context of theoretical framework. It also helped ensure that the item pool did not include items that load on additional factors and represent constructs beyond the boundaries of the conceptual framework. Following refinements based on the study outcomes, the item pool was reduced to 31 items across the four domains. The correlations between factors ranged from weak to strong, indicating nuanced relationship between variables.

Subsequently, CFA was used to test the fit of the data to the conceptual model in which maladaptive overcontrol was conceptualised as a higher-order factor and the four deficits were conceptualised as the lower-order factors. The scale was also further refined based on standardised loadings and modification indices. The aim of the modifications was to increase model fit without compromising the theoretical validity of the scale (Hair et al., 2019). Following these refinements, the final item pool was reduced to 26 items across the four domains. The results of the study indicated good fit of the higher-order model, further validating the results of the EFA, and providing preliminary evidence validating Lynch's (2018) theoretical framework. Internal consistency of the scale was also evaluated at this stage, yielding satisfactory results. Satisfactory model fit indicated that the relationships between the four deficits can be explained by a higher-order construct of maladaptive overcontrol to a satisfactory extent. However, the study did reveal that there was some covariance between the domains unexplained by the higher-order factor of maladaptive overcontrol, suggesting that once more research on the theory emerges, modifications to the model might be necessary.

Collectively, the factor-analytic studies provided preliminary construct validity evidence. The methodological rigour, demonstrated by adherence to best practice recommendations from wide range of psychometric sources, underscores the credibility and reliability of the findings. The studies also revealed some interesting nuances regarding the relationships between the domains. Researchers ought to be mindful of these nuances when designing, conducting, and interpreting future studies within the Lynch's (2018) theoretical framework.

8.1.6 Psychometric properties of the OAQ and the links between maladaptive overcontrol and other psychological constructs.

The last study aimed to further evaluate its psychometric properties. The results of the study indicated significant, positive correlations across the OAQ and its subscales, indicating meaningful connections between the four deficits and the overarching construct of maladaptive overcontrol, and cross-validating the results of the factor-analytic studies. Further, satisfactory 4-week test-retest reliability was found, signifying stability of the measure over time.

Further, the study aimed to place maladaptive overcontrol within wider psychological literature by exploring its relationships with several other psychological constructs and theories relevant to Lynch's (2018) theorising. The findings revealed a complex landscape of correlations. It was found that individuals reporting higher levels of maladaptive overcontrolled deficits presented with a more sensitive behavioural inhibition system and lower ego-resilience, corroborating some key elements of Lynch's (2018) theorising regarding avoidance of novelty, threat sensitivity, and decreased ability to adjust one's level of self-control to changing circumstances. Some convergence was found between Lynch's conceptualisation of maladaptive overcontrol and ego-overcontrol (Block, 1971; Block & Block, 1980), with the findings underscoring Lynch's framework as more comprehensive than the framework of ego-control. Results pertaining to behavioural activation were mixed. Highly overcontrolled individuals were found to be less reward-responsive in socioemotional contexts and less likely to seek out new and exciting activities in everyday life. This in accordance with the neurobiosocial theory. In relation one's drive to achieve desired goals, the results indicated that those with profound deficits in receptivity and openness are less likely to take risks and ruthlessly and selfishly pursue their desired goals, while those with a stronger preference for structure, order, and predictability, and more deficits in social connectedness and emotional expressivity are more motivated to pursue their goals in a detached, task-oriented manner. The study yielded mixed results in relation to other theories of self-control. It appeared that the considered frameworks designed to measure low self-control did not consistently measure a construct that was different to maladaptive overcontrol, as conceptualised by the OAQ. This warrants further investigation to better understand the differences between undercontrol, overcontrol, and optimal self-control. A complicated pattern of correlations was also

revealed between the dimensions of maladaptive overcontrol, as measured by the OAQ, and impulsive behaviour, understood through the lenses of the UPPS-P framework. The results did not fully align with the hypotheses, suggesting that the relationship between maladaptive overcontrol and this popular conceptualisation of impulsive behaviour may be more complex than expected. The results did, however, highlight self-control as a multi-faceted construct, and further underscored that the linear view is limiting. As such, maladaptive overcontrol in relation to this and other existing models of impulsivity ought to be further explored.

8.1.7 Research outcomes.

The results of the studies conducted as part of this research programme present the OAQ as a valid and reliable measure that can be confidently used in English-speaking general population research to investigate the construct of maladaptive overcontrol. Given the theory-based scale development process, the results of the studies also provide preliminary support for several elements of Lynch's (2018) Neurobiosocial Theory for Disorders for Overcontrol. The new scale provides a solid foundation for future studies to be able to explore the proposed development, maintenance, and presentations of maladaptive overcontrol. Importantly, the findings from the project also highlight the need to move away from the erroneous linear view of self-control and to embrace the quadratic view which recognises that both excessive and insufficient self-control can be maladaptive.

8.2 Sample considerations.

Given limited resources, while maintaining rigorous research standards, the recruitment strategy applied throughout this project also prioritised cost- and time-efficiency. The recruitment for the studies used to develop and validate the OAQ was limited to English-speaking individuals from the general population. The participants in the larger-scale quantitative studies described in Chapter 6 and Chapter 7 participants were recruited exclusively through the Prolific platform. This recruitment strategy proved successful in achieving the objective of a project and producing a measure; it does, however, present certain limitations. Some of the study-specific recruitment issues have been discussed in detail in the respective chapters (e.g., issues associated with using ASC-WP as an overcontrol screening measure and issues associated with online, social media-based recruitment in Chapter 5). The current

section provides a more general discussion of the constraints associated with the recruitment strategy and considers the generalisability and utility of the OAQ in its current form.

8.2.1 General population and cultural and linguistic considerations.

Given the time constraints of this research and the extended time typically required for a researcher to be able to access clinical and forensic populations, it was decided to conduct the initial development and validation of the OAQ within the general population. Conducting research in clinical and forensic settings involves complex ethical and logistical challenges (see Shapiro et al., 2013) that would not have been feasible within the timeframe of this PhD programme. Since the primary aim of the project was to develop a measure that could be used to identify maladaptively overcontrolled individuals for research purposes, recruiting a general population sample for the initial development and validation of the scale was deemed sufficient, and allowed for a timely and efficient conduct of the project.

However, this decision does limit the generalisability and utility of the OAQ in its current form to non-clinical, non-forensic samples (see Briere & Elliott, 1998). While general population studies can provide valuable insights about the nature of overcontrol and the mechanisms behind the development and maintenance of maladaptive overcontrolled coping, the prevalence, severity, and behavioural manifestations of maladaptive overcontrol are likely to differ across various clinical and forensic populations (Clark & Watson, 2016). General population studies may not capture the full spectrum of overcontrolled traits, behaviours, and difficulties that would be present in populations characterised by more severe levels of psychopathology (Clark & Watson, 2016, 2019). Different populations are likely to require different cut-off scores, and certain OAQ items may need to be revised to better capture population-specific presentations of maladaptive overcontrol if identified in future research. To extend the utility of the scale to clinical and forensic settings, validation studies within these specific samples are needed (Clark & Watson, 2019). Further, nothing is known about the OAQ's predictive validity – a type of validity concerned with whether the scores on the scale can accurately predict future outcomes (Colton & Covert, 2007). The ability of a scale to predict future outcomes is often crucial in clinical and forensic settings in relation to risk assessment and treatment

planning and evaluation. As such, future studies should also test predictive validity of the OAQ in these contexts.

Despite the OAQ having been developed primarily for research purposes and in general population, several strategies applied during the process may facilitate its future validation for use in clinical and forensic settings. Firstly, the measure is based on a theoretical framework that specifically relates to overcontrolled psychopathology. As such, the scale's items and dimensions are also directly relevant to clinical and forensic contexts where maladaptive overcontrol has been shown to be prevalent, with the measure designed with behaviours, patterns, and mental states pertinent to psychopathology. Secondly, the content validity of the scale has been assessed by both researchers and clinical professionals with experience of treating overcontrolled clients. This approach ensures that the items of the scale are relevant both theoretically and practically, with the involvement of clinical professionals increasing the likelihood that the items are meaningful also in clinical contexts. These strategies provide a robust foundation for future validation work in clinical and forensic contexts. Future studies should focus on building on this foundation by testing the psychometric properties of the OAQ in various clinical and forensic contexts, developing population-specific clinical cut-off scores, and evaluating the scale's utility in predicting future outcomes and informing client care and treatment.

Lastly, the studies presented within this thesis have been conducted exclusively with English-speaking samples from a limited number of countries. While the initial item generation study and the pre-testing did not apply specific requirements as to the country of residence of participants who can take part, most participants within the samples still resided in the UK or the USA. As such, it was necessary to recognise that the new scale inevitably carried the influence of the social, cultural, and economic characteristics of participants from these backgrounds, and the participation in the last two studies was limited to English-speaking WEIRD countries. The aim of this was to strike a balance between sensitivity, specificity, and methodological rigour necessary to create a valid and reliable instrument. However, to extend the utility of the OAQ to non-English speaking populations and other cultures, linguistic adaptation of the questionnaire must be carried out, and multicultural validation studies must be

conducted to establish how well the scale measures the latent construct across various cultures, with the item pool adjusted if and as necessary.

8.2.3 Prolific recruitment.

The samples for the studies presented in Chapters 6 and 7 were recruited exclusively from the Prolific participant pool. Prolific is an online crowdsourcing platform for academic and market research. Use of such online data collection platforms – including Prolific, Mechanical Turk (MTurk), and CrowdFlower – has increased substantially in the recent years. This is largely because the platforms offer time- and cost-efficient solution for recruiting substantial participant samples that meet specific study criteria (Palan & Schitter, 2018).

This recruitment strategy does carry certain limitations. For example, while online crowdsourcing platforms may offer access to a wider demographic and decreased likelihood of AI-generated data compared to other recruitment methods, the generalizability of the results remains somewhat limited. This is because online crowdsourcing samples are limited to technologically active, self-selecting samples of participants who are motivated by monetary rewards (Palan & Schitter, 2018; Peer et al., 2017). Additionally, the ethnic diversity of participants on some of the platforms, including Prolific, has been indicated to be relatively low (Peer et al., 2019). Further, previous research on online crowdsourcing platforms has raised some concerns over the quality of the data collected on such platforms. This was, however, particularly in the context of MTurk. Some studies indicated that MTurk recruitment generally produces good quality data (e.g., Kees et al., 2017; Smith et al., 2016), while others reported problematic respondent behaviours likely to negatively impact the data quality (e.g., Barends & De Vries, 2019; Kan & Drummey, 2018; Necka et al., 2016).

The concerns over data quality were not as substantial on Prolific when looking at alternative crowdsourcing platforms. A comparative study by Peer and colleagues (2017) explored the quality of data obtained through MTurk, Prolific, and CrowdFlower based on the time spent completing the questionnaires, accuracy of responses to attention check questions, number of studies that the platform participants take part in, and the reliability of established psychometric measures included in the studies. The results of the study underscored Prolific as a platform superior to both

MTurk and CrowdFlower, with Prolific participants taking fewer studies on average, failing less attention checks, and having lower levels of dishonest behaviour. Prolific was also found to provide the higher quality data compared to MTurk in studies by Eyal and colleagues (2021) and Douglas and colleagues (2023), including relative to the cost when compared to MTurk (Douglas et al., 2023). In addition, a substantial benefit of using Prolific for online participant recruitment is that the platform thoroughly verifies their participants on sign-up, further assuring higher data quality, and decreasing the likelihood of AI-assisted and bot-like responses (Peer et al., 2021). It also offers researchers a chance to pre-screen for participants not only by specific demographic criteria, but also for high study approval rates, which is now standard practice in the social sciences (Douglas et al., 2023). While inattentive and rushed responding can still occur, attention check questions are a requirement in all studies conducted on Prolific to help mitigate this risk and ensure reliable data. Moreover, monetary rewards of participants on Prolific (min. \$8 per hour *pro rata*) are substantially higher than those on MTurk (min. \$0.01 per assignment), with MTurk's low compensation of workers previously raising ethical concerns (Hara et al., 2018).

Despite potential drawbacks, the time- and cost-efficiency of online crowdsourcing platforms made the recruitment strategy the most viable option for the quantitative studies that required relatively large participant samples. Prolific was chosen due to previous studies indicating the highest quality data compared to the cost and extensive participant verification efforts. The studies presented in the current thesis yielded no obvious concerns as to the quality of data collected on Prolific. The results of the EFA and CFA studies yielded similar results as to item performance despite being conducted using different samples of participants, supporting the reliability of the findings through cross-validation. Further, the study indicated satisfactory internal consistency reliability of most scales and subscales within the battery of scales, were indicated to have good internal consistency reliability, supporting the adequate quality of the collected data. The retention of participants for the 4-week test-retest reliability follow-up was also satisfactory. Nevertheless, it is strongly recommended that future studies explore the psychometric properties of the OAQ also in other participant samples.

8.3 Future directions: scale and theory refinement.

Building on the findings from this programme of research, it is evident that both the OAQ and the framework need further refinement. In this section, two key recommendations for improvement are discussed.

Firstly, the LFC subscale requires further validation and refinement to ensure that it accurately reflects maladaptive behaviours characteristic of highly overcontrolled individuals. Its relatively low correlations between with the other three subscales and the patterns of correlations with other constructs found in the last study indicate that the LFC subscale may be measuring behaviours that could be both adaptive and maladaptive – depending on the context in which they occur. Behaviours represented by the subscale items, such as careful planning and preparation, can be adaptive when the situational context demands it. However, in the context of overcontrol, they can be maladaptive when the environmental demands call for a reduction in self-control. The dual nature of the LFC subscale in its current form presents a potential risk of a jingle fallacy where adaptive and maladaptive aspects could be conflated. This mirrors the distinction between constructs like functional and dysfunctional impulsivity (see Dickman, 1990), where impulsivity can be adaptive in some contexts (e.g., when quick decision-making is required) but maladaptive when inhibition of impulses is called for. Similarly, it is important to distinguish between a temporary lack of motivation and avolition, which represents a more pervasive and dysfunctional disengagement from and disinterest in goal-directed behaviour (see e.g., DeRosse et al., 2019; Raffard et al., 2022). Therefore, future work should focus on refining the LFC subscale to ensure a clearer differentiation between behaviours that indicate maladaptive inflexible overcontrol and those associated with optimal or flexible control.

Secondly, the CFA results indicated unaccounted-for covariance between the LFC and LRO domains, as well as the LSC and PIE domains. This suggests that the conceptualisation of the four core deficits may require updating. One possibility is that the LFC and the LRO domains and the LSC and the PIE domains form two dyads that are more closely conceptually related when compared to other deficits. Considering how the deficits are currently defined within the theory, it is plausible that the LFC x LRO dyad reflects broadly defined psychological inflexibility, while the LSC x PIE

dyad represents dysfunctional socio-emotional functioning. Further, there could exist a causal relationship between the two dyads. It could be that maladaptive behaviours reflected by the LFC and LRO dyad *result in* socio-emotional difficulties reflected by the LSC and PIE dyad. This is plausible given that some of the socio-emotional experiences reflected by the LSC and PIE dyad – such as loneliness and feeling misunderstood or like an outsider – could align both with under- and overcontrolled mental health conditions. Nevertheless, this remains speculative, and further empirical research is needed to better understand the relationships between the deficits and refine Lynch's (2018) theoretical framework.

8.4 Embracing the iterative approach to scale and theory development.

To date, the absence of a scale measuring maladaptive overcontrol has been hindering progress in testing and applying the theory in practice. The work conducted during this PhD programme marks the first successful attempt at developing a self-report scale to assess maladaptive overcontrol in adults, seen through the lens of Lynch's (2018) Neurobiosocial Theory for Disorders of Overcontrol. Now, with the first version of the scale ready to be used in future research, academics and clinicians can begin building an empirical base for maladaptive overcontrol.

However, a successful initial attempt at developing the scale does not imply that the work is complete. Rather, it marks only the beginning. In fact, both the scale and the theoretical framework underpinning the scale are novel and – as evident from the outputs of the research in this thesis and the previous paragraph – will require continuous revisions to align with the evolving evidence. Results from the last two studies indicated that both the OAQ and Lynch's (2018) theoretical framework may have certain limitations (e.g., the unaccounted covariance between the OAQ's dimensions and Lynch's conceptualisation of overcontrol and undercontrol as mutually exclusive). Because scale and theory development are iterative processes that inform each other, revising the scale while neglecting the potential imperfections of the theory would be a mistake. Paying equal attention to improving both the scale and the theory ensures that the theory appropriately reflects problematic overcontrolled patterns while the scale remains theoretically relevant.

In the words of Clark and Watson (2019, p. 9):

“Good scale construction is an iterative process involving an initial cycle of preliminary measure development, data collection, and psychometric evaluation, followed by at least one additional cycle of revision of both measure and construct, data collection, psychometric evaluation, revision, and so forth. The most often neglected aspect of this process is revision of the target construct’s conceptualization. Too often, scale developers assume that their initial conceptualization is entirely correct, considering only the measure as open to revision. However, it is critical to remain open to rethinking one’s initial construct – to ‘listen to the data’ not ‘make the data talk.’”

Future research should embrace the iterative approach to ensure that the construct of maladaptive overcontrol remains relevant over time. This proactive approach will help prevent the novel theoretical framework being overshadowed by other research, alike its predecessors, as well as ensure that individuals who struggle with maladaptive overcontrol receive the attention and support that they deserve.

8.5 Navigating the complexities of psychometric research.

While psychometric measures are considered essential to researching human psychology, creating valid and reliable scales that accurately capture the underlying theoretical constructs is far from straightforward. Especially for early career researchers without prior experience in scale development, navigating psychometric literature can feel like venturing through a dense jungle of conflicting information.

Introductory literature aimed at novices only provides selected information and often contains little detail and nuance regarding applying different statistical methods at different stages of scale development. More advanced information is scattered across an array of empirical papers written in a specialist language, making accessing and understanding the necessary information rather challenging. Despite database searches yielding thousands of results, there are few papers written in a more accessible language that provide methodological insights into the less frequently spoken, yet critical, aspects of the psychometric process and analysis. The difficulty to find good quality psychometric sources in this complex landscape, in turn, leads to overreliance on a few novice-friendly papers. Further, an overwhelming majority of the existing

Likert-type scale development and evaluation studies employs less-than-optimal methods, including use of arbitrary methods developed based on continuous data and suboptimal for use with ordinal data (e.g., use of the maximum likelihood estimation method as default, and inappropriately interpreting Hu and Bentler's [1999] model fit cut-off scores and the Cronbach's [1951] alpha statistic). This makes it challenging for a trainee researcher to discern best practices without the continuous guidance and support of experienced psychometricians.

Addressing the above challenges is critical to improve the rigour of psychometric studies and facilitate the application of best practices on a wider scale. Collaborative efforts of novice scale developers and experienced psychometricians are needed to create accessible and comprehensive guides that will make it easier to navigate the complex landscape of psychometric scale development and evaluation.

8.6 Conclusion

To conclude, this research programme made an original contribution to social science by developing a much-needed self-control measure designed to assess individuals for maladaptive overcontrol – the Overcontrol Assessment Questionnaire, or the OAQ. The scale development and validation were theory-driven, guided by the Neurobiosocial Theory for Disorders of Overcontrol. The procedures utilised were rigorous, guided by an extensive review of psychometric literature. The results provided encouraging preliminary content and construct validity evidence, and indicated satisfactory model fit, internal consistency and composite reliability, as well as 4-week test-retest reliability. Together, they also provided preliminary support for underpinning the theoretical model of maladaptive overcontrol. As such, the scale can be confidently applied in future research. The final version of the OAQ, alongside participant instructions and scoring procedures, is provided in Appendix M.

Future studies should aim to extend the understanding the nature of maladaptive overcontrol – the mechanisms behind its development, the factors that contribute to its reinforcement and maintenance, the presentations of maladaptive overcontrol across different populations, and the relationships between maladaptive overcontrol and other psychological constructs and theories, particularly those concerning personality and self-control. The researchers should also aim to obtain

clinical and multi-cultural validity of the scale to extend the utility of the scale. The links to psychopathology and offending behaviour, as well as treatment avenues and treatment outcomes, also ought to be explored in more detail, so that tailored help can be offered to problematically overcontrolled individuals on a wider scale. Further, in the light of the outcomes of this research, as well as when new evidence emerges, both the theoretical framework and the scale ought to be refined.

Lastly, the arguments and findings of the studies presented in the current thesis clearly highlight that it is necessary to move away from the linear view of self-control. It is now clear that both excessive and insufficient self-control can be problematic and lead to adverse life outcomes. Going forward, neither overcontrolled nor undercontrolled issues can be overlooked by researchers and practitioners. It is crucial that the similarities and differences in manifestations between and the mechanisms behind maladaptive overcontrol and undercontrol are extensively investigated, so that individuals who struggle with either of these can be better understood and able to access appropriate, tailored help.

References

- Abdelazeem, B., Hamdallah, A., Rizk, M. A., Abbas, K. S., El-Shahat, N. A., Manasrah, N., Mostafa, M. R., & Eltobgy, M. (2023). Does usage of monetary incentive impact the involvement in surveys? A systematic review and meta-analysis of 46 randomized controlled trials. *PloS One*, *18*(1), Article e0279128. <https://doi.org/10.1371/journal.pone.0279128>
- Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs: General and Applied*, *80*(7), 1–37. <https://doi.org/10.1037/h0093906>
- Achenbach, T. M., Ivanova, M. Y., Rescorla, L. A., Turner, L. V., & Althoff, R. R. (2016). Internalizing/externalizing problems: Review and recommendations for clinical and research applications. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(8), 647–656. <https://doi.org/10.1016/j.jaac.2016.05.012>
- Alloy, L. B., Abramson, L. Y., Walshaw, P. D., Gerstein, R. K., Keyser, J. D., Whitehouse, W. G., Urosevic, S., Nusslock, R., Hogan, M. E., & Harmon-Jones, E. (2009). Behavioral approach system (BAS)–relevant cognitive styles and bipolar spectrum disorders: Concurrent and prospective associations. *Journal of Abnormal Psychology*, *118*(3), 459–471. <https://doi.org/10.1037/a0016604>
- Allport, G. W. (1961). *Pattern and growth in personality*. Holt, Reinhart & Winston.
- Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social & Administrative Pharmacy: RSAP*, *15*(2), 214–221. <https://doi.org/10.1016/j.sapharm.2018.03.066>
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*(2), 356–366. <https://doi.org/10.1037/0022-3514.86.2.356>
- Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis of latent variable models using Mplus*. Retrieved Aug 19, 2024, from <http://www.statmodel.com/download/BayesAdvantages18.pdf>

- Barclay, J. E., & Weaver, H. B. (1962). Comparative reliabilities and ease of construction of Thurstone and Likert Attitude Scales. *Journal of Social Psychology*, 58(1), 109–120.
- Barends, A. J., & De Vries, R. E. (2019). Noncompliant responding: Comparing exclusion criteria in MTurk personality research to improve data quality. *Personality and Individual Differences*, 143, 84–89. <https://doi.org/10.1016/j.paid.2019.02.015>
- Barnes, S. J., & Pinel, J. P. (2001). Conditioned effects of kindling. *Neuroscience and Biobehavioral Reviews*, 25(7–8), 745–751. [https://doi.org/10.1016/s0149-7634\(01\)00054-9](https://doi.org/10.1016/s0149-7634(01)00054-9)
- Bartlett, M. S. (1951). The effect of standardization on a χ^2 approximation in factor analysis. *Biometrika*, 38(3/4), 337–344. <https://doi.org/10.2307/2332580>
- Baudinet, J., Simic, M., Griffiths, H., Donnelly, C., Stewart, C., & Goddard, E. (2020). Targeting maladaptive overcontrol with Radically Open Dialectical Behaviour Therapy in a day programme for adolescents with restrictive eating disorders: An uncontrolled case series. *Journal of Eating Disorders*, 8(1), 68. <https://doi.org/10.1186/s40337-020-00338-9>
- Baudinet, J., Stewart, C., Bennett, E., Konstantellou, A., Parham, R., Smith, K., Hunt, K., Eisler, I., & Simic, M. (2021). Radically Open Dialectical Behaviour Therapy adapted for adolescents: A case series. *BMC Psychiatry*, 21, Article 462. <https://doi.org/10.1186/s12888-021-03460-3>
- Baudinet, J., Watson, C., Brothwood, P. L., Parham, R., Smith, L., Snowden, N., Konstantellou, A., Hunt, K., & Simic, M. (2022). Adolescent experience of radically open dialectical behaviour therapy: A qualitative study. *BMC Psychiatry*, 22, Article 466. <https://doi.org/10.1186/s12888-022-04114-8>
- Bauer, R. M. (2007). Evidence-based practice in psychology: Implication for research and research training. *Journal of Clinical Psychology*, 63(7), 685–694. <https://doi.org/10.1002/jclp.20374>
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74(5), 1252–1265. <https://doi.org/10.1037/0022-3514.74.5.1252>
- Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1994). *Losing control: How and why people fail at self-regulation*. Academic Press.

- Baumeister, R. F., Tice, D. M., & Vohs, K. D. (2018). The Strength Model of Self-Regulation: Conclusions from the second decade of willpower research. *Perspectives on Psychological Science*, 13(2), 141–145.
<https://doi.org/10.1177/1745691617716946>
- Baumeister, R. F., & Vohs, K. D. (2018). Strength model of self-regulation as limited resource: Assessment, controversies, update. In J. M. Olson, & M. P. Zanna (Eds.), *Advances in experimental social psychology* (pp. 67–127). Elsevier Academic Press. <https://doi.org/10.1016/bs.aesp.2016.04.001>
- Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The Strength Model of Self-Control. *Current Directions in Psychological Science*, 16(6), 351–355.
<https://doi.org/10.1111/j.1467-8721.2007.00534.x>
- Baumeister, R. F., Wright, B. R. E., & Carreon, D. (2019). Self-control “in the wild”: Experience sampling study of trait and state self-regulation. *Self and Identity*, 18(5), 494–528. <https://doi.org/10.1080/15298868.2018.1478324>
- Beauducel, A., & Herzberg, P. Y. (2006). On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural Equation Modeling*, 13(2), 186–203.
https://doi.org/10.1207/s15328007sem1302_2
- Beck, K. (2020). Ensuring Content Validity of Psychological and Educational Tests—The Role of Experts. *Frontline Learning Research*, 8(6), 1–37.
- Billieux, J., Gay, P., Rochat, L., Khazaal, Y., Zullino, D., & Van der Linden, M. (2010). Lack of inhibitory control predicts cigarette smoking dependence: Evidence from a non-deprived sample of light to moderate smokers. *Drug and Alcohol Dependence*, 112(1), 164–167.
<https://doi.org/10.1016/j.drugalcdep.2010.06.006>
- Blackford, J. U., Clauss, J. A., & Benningfield, M. M. (2018). The neurobiology of behavioral inhibition as a developmental mechanism. In K. Pérez-Edgar & N. A. Fox (Eds.), *Behavioral inhibition: Integrating theory, research, and clinical perspectives* (pp. 113–134). Springer International Publishing.
https://doi.org/10.1007/978-3-319-98077-5_6
- Blair, J., & Conrad, F. G. (2011). Sample size for cognitive interview pretesting. *Public Opinion Quarterly*, 75(4), 636–658.
<https://doi.org/10.1093/poq/nfr035>
- Bleidorn, W., Schwaba, T., Zheng, A., Hopwood, C. J., Sosa, S. S., Roberts, B. W., & Briley, D. A. (2022). Personality stability and change: A meta-analysis of

- longitudinal studies. *Psychological Bulletin*, 148(7–8), 588–619.
<https://doi.org/10.1037/bul0000365>
- Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research*. Charles C Thomas Publisher. <https://doi.org/10.1037/13141-000>
- Block, J. (1971). *Lives through time*. Bancroft Book.
- Block, J. (1993). *Ego-resilience through time* [ERIC Document Reproduction Service No. ED356879].
- Block, J. H., & Block, J. (1980). The Role of Ego-Control and Ego-Resiliency in the Organization of Behavior. In W. A. Collins (Ed.), *Development of cognition, affect, and social relations* (pp. 39-101). Psychology Press.
- Block, J., & Kremen, A. M. (1996). IQ and ego-resiliency: conceptual and empirical connections and separateness. *Journal of Personality and Social Psychology*, 70(2), 349-361. <https://doi.org/10.1037/0022-3514.70.2.349>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Bohane, L., Maguire, N., & Richardson, T. (2017). Resilients, overcontrollers and undercontrollers: A systematic review of the utility of a personality typology method in understanding adult mental health problems. *Clinical Psychology Review*, 57, 75–92. <https://doi.org/10.1016/j.cpr.2017.07.005>
- Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., Waltz, T., & Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, 42(4), 676–688. <https://doi.org/10.1016/j.beth.2011.03.007>
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological review*, 111(4), 1061-1071. <https://doi.org/10.1037/0033-295X.111.4.1061>
- Bowling, N. A., Gibson, A. M., & DeSimone, J. A. (2022). Stop with the questions already! Does data quality suffer for scales positioned near the end of a lengthy questionnaire? *Journal of Business and Psychology*, 37(5), 1099-1116. <https://doi.org/10.1007/s10869-021-09787-8>

- Briere, J., & Elliott, D. M. (1998). Clinical utility of the Impact of Event Scale: Psychometrics in the general population. *Assessment*, 5(2), 171–180.
<https://doi.org/10.1177/107319119800500207>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Press.
- Buker, H. (2011). Formation of self-control: Gottfredson and Hirschi's general theory of crime and beyond. *Aggression and Violent Behavior*, 16(3), 265–276.
<https://doi.org/10.1016/j.avb.2011.03.005>
- Butler, E. A., Egloff, B., Wilhelm, F. H., Smith, N. C., Erickson, E. A., & Gross, J. J. (2003). The social consequences of expressive suppression. *Emotion*, 3(1), 48–67. <https://doi.org/10.1037/1528-3542.3.1.48>
- Caldwell, J. G., & Shaver, P. R. (2012). Exploring the cognitive-emotional pathways between adult attachment and ego-resiliency. *Individual Differences Research*, 10(3), 141–152.
- Carver, C. S. (2004). Negative affects deriving from the behavioral approach system. *Emotion*, 4(1), 3–22. <https://doi.org/10.1037/1528-3542.4.1.3>
- Carver, C. S. (2009). Threat Sensitivity, Incentive Sensitivity, and the Experience of Relief. *Journal of Personality*, 77(1), 125–138.
<https://doi.org/10.1111/j.1467-6494.2008.00540.x>
- Carver, C. S., & Scheier, M. F. (1981). The self-attention-induced feedback loop and social facilitation. *Journal of Experimental Social Psychology*, 17(6), 545–568. [https://doi.org/10.1016/0022-1031\(81\)90039-1](https://doi.org/10.1016/0022-1031(81)90039-1)
- Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin*, 92(1), 111–135. <https://doi.org/10.1037/0033-2909.92.1.111>
- Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, 67(2), 319–333. <https://doi.org/10.1037/0022-3514.67.2.319>
- Causadias, J. M., Salvatore, J. E., & Sroufe, L. A. (2012). Early patterns of self-regulation as risk and promotive factors in development: A longitudinal study from childhood to adulthood in a high-risk sample. *International Journal of Behavioral Development*, 36(4), 293–302.
<https://doi.org/10.1177/0165025412444076>

- Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement, 18*(3), 205–215. <https://doi.org/10.1177/014662169401800302>
- Choate, A. M., Fatimah, H., & Bornovalova, M. A. (2021). Comorbidity in borderline personality: Understanding dynamics in development. *Current Opinion in Psychology, 37*, 104–108. <https://doi.org/10.1016/j.copsyc.2020.09.015>
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. <https://doi.org/10.1037/1040-3590.7.3.309>
- Clark, L. A., & Watson, D. (2016). Constructing validity: Basic issues in objective scale development. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (4th ed., pp. 187–203). American Psychological Association. <https://doi.org/10.1037/14805-012>
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31*(12), 1412–1427. <https://doi.org/10.1037/pas0000626>
- Cobb-Clark, D. A., Dahmann, S. C., Kamhöfer, D. A., & Schildberg-Hörisch, H. (2022). The Predictive Power of Self-Control for Life Outcomes. *Journal of Economic Behavior & Organization, 197*, 725–744. <https://doi.org/10.1016/j.jebo.2022.02.028>
- Colton, D., & Covert, R. W. (2007). *Designing and Constructing Instruments for Social Research and Evaluation*. John Wiley & Sons.
- Connor, D. E., Edwards, G., Fletcher, K. E., Baird, J., Barkley, R. A., & Steingard, R. J. (2003). Correlates of comorbid psychopathology in children with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*(2), 193–200. <https://doi.org/10.1097/00004583-200302000-00013>
- Conrad, F., & Blair, J. (1996). From impressions to data: Increasing the objectivity of cognitive interviews. *Proceedings of the section on survey research methods, annual meetings of the American Statistical Association*, pp. 1–10.
- Conrad, F., Blair, J., & Tracy, E. (1999). Verbal reports are data! A theoretical approach to cognitive interviews. *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, pp. 11–20.
- Converse, P. D., Beverage, M. S., Vaghef, K., & Moore, L. S. (2018). Self-control over time: Implications for work, relationship, and well-being outcomes.

- Journal of Research in Personality*, 73, 82–92.
<https://doi.org/10.1016/j.jrp.2017.11.002>
- Cornejo, B., Vela, B., Vilca, L. W., Vallejos, M., & Huancahuire-Vega, S. (2021). Psychometric properties of the COVID-19 protective motivation scale in peruvians during the health emergency. *Journal of Primary Care & Community Health*, 12, Article 21501327211051935.
<https://doi.org/10.1177/21501327211051935>
- Cornwall, P. L., Simpson, S., Gibbs, C., & Morfee, V. (2021). Evaluation of Radically Open Dialectical Behaviour Therapy in an adult community mental health team: Effectiveness in people with autism spectrum disorders. *BJPsych Bulletin*, 45(3), 146–153. <https://doi.org/10.1192/bjb.2020.113>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104.
<https://doi.org/10.1037/0021-9010.78.1.98>
- Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, 13(6), 653–665. [https://doi.org/10.1016/0191-8869\(92\)90236-I](https://doi.org/10.1016/0191-8869(92)90236-I)
- Costello, A., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10, 1–9.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. <https://doi.org/10.1007/BF02310555>
- Cumberland-Li, A., Eisenberg, N., & Reiser, M. (2004). Relations of young children's agreeableness and resiliency to effortful control and impulsivity. *Social Development*, 13(2), 193–212. <https://doi.org/10.1111/j.1467-9507.2004.000263.x>
- Cyders, M. A., & Smith, G. T. (2007). Mood-based rash action and its components: Positive and negative urgency. *Personality and Individual Differences*, 43(4), 839–850. <https://doi.org/10.1016/j.paid.2007.02.008>
- Cyders, M. A., Smith, G. T., Spillane, N. S., Fischer, S., Annus, A. M., & Peterson, C. (2007). Integration of impulsivity and positive mood to predict risky behavior: Development and validation of a measure of positive urgency. *Psychological Assessment*, 19(1), 107–118. <https://doi.org/10.1037/1040-3590.19.1.107>

- D'Silva, K., & Duggan, C. (2010). Revisiting the overcontrolled–undercontrolled typology of violent offenders. *Personality and Mental Health*, 4(3), 193–205. <https://doi.org/10.1002/pmh.130>
- Dalal, D. K., Carter, N. T., & Lake, C. J. (2014). Middle response scale options are inappropriate for ideal point scales. *Journal of Business and Psychology*, 29(3), 463–478. <https://doi.org/10.1007/s10869-013-9326-5>
- Daly, M., Delaney, L., Egan, M., & Baumeister, R. F. (2015). Childhood self-control and unemployment throughout the life span: Evidence from two British cohort studies. *Psychological Science*, 26(6), 709–723. <https://doi.org/10.1177/0956797615569001>
- Davey, L., Day, A., & Howells, K. (2005). Anger, over-control and serious violent offending. *Aggression and Violent Behavior*, 10(5), 624–635. <https://doi.org/10.1016/j.avb.2004.12.002>
- Davidson, R. J. (1992). Emotion and affective style: Hemispheric substrates. *Psychological Science*, 3(1), 39–43. <https://doi.org/10.1111/j.1467-9280.1992.tb00254.x>
- Davidson, R. J. (1998). Affective style and affective disorders: Perspectives from affective neuroscience. *Cognition and Emotion*, 12(3), 307–330. <https://doi.org/10.1080/026999398379628>
- Davisson, E. K., & Hoyle, R. H. (2017). Varieties of self-control and their personality correlates. In *Handbook of Self-Regulation, Third Edition: Research, Theory, and Applications* (3rd ed., pp. 396–413). Guilford Press.
- De Ridder, D., Kroese, F., & Gillebaart, M. (2018). Whatever happened to self-control? A proposal for integrating notions from trait self-control studies into state self-control research. *Motivation Science*, 4(1), 39–49. <https://doi.org/10.1037/mot0000062>
- De Ridder, D. T. D., De Boer, B. J., Lugtig, P., Bakker, A. B., & van Hooft, E. A. J. (2011). Not doing bad things is not equivalent to doing the right thing: Distinguishing between inhibitory and initiatory self-control. *Personality and Individual Differences*, 50(7), 1006–1011. <https://doi.org/10.1016/j.paid.2011.01.015>
- De Ridder, D. T. D., Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M., & Baumeister, R. F. (2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Personality and*

- Social Psychology Review*, 16(1), 76–99.
<https://doi.org/10.1177/1088868311418749>
- De Vries, R. E., Realo, A., & Allik, J. (2016). Using personality item characteristics to predict single-item internal reliability, retest reliability, and self–other agreement. *European Journal of Personality*, 30(6), 618–636.
<https://doi.org/10.1002/per.2083>
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale Development: Theory and Applications*. SAGE Publications.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. <https://doi.org/10.1176/appi.books.9780890425596>
- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, 64(1), 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dinno, A. (2009). Exploring the sensitivity of Horn’s parallel analysis to the distributional form of random data. *Multivariate Behavioral Research*, 44(3), 362–388. <https://doi.org/10.1080/00273170902938969>
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30, 412–431. [https://doi.org/10.1016/0001-6918\(69\)90065-1](https://doi.org/10.1016/0001-6918(69)90065-1)
- Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE*, 18(3), e0279720.
<https://doi.org/10.1371/journal.pone.0279720>
- Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3(4), 465–476. <https://doi.org/10.1111/j.1754-9434.2010.01273.x>
- Drennan, J. (2003). Cognitive interviewing: Verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing*, 42(1), 57–63.
<https://doi.org/10.1046/j.1365-2648.2003.02579.x>
- Du Toit, L., & Duckitt, J. (1990). Psychological Characteristics of Over- and Undercontrolled Violent Offenders. *Journal of Psychology*, 124(2), 125–141.
<https://doi.org/10.1080/00223980.1990.10543210>
- Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, 45(3), 259–268.
<https://doi.org/10.1016/j.jrp.2011.02.004>

- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Eaton, N. R., Krueger, R. F., Keyes, K. M., Skodol, A. E., Markon, K. E., Grant, B. F., & Hasin, D. S. (2011). Borderline personality disorder co-morbidity: Relationship to the internalizing-externalizing structure of common mental disorders. *Psychological Medicine*, 41(5), 1041–1050. <https://doi.org/10.1017/S0033291710001662>
- Edwards, A. L., & Kenney, K. C. (1946). A comparison of the Thurstone and Likert techniques of attitude scale construction. *Journal of Applied Psychology*, 30(1), 72–83. <https://doi.org/10.1037/h0062418>
- Eggum-Wilkens, N. D., Reichenberg, R. E., Eisenberg, N., & Spinrad, T. L. (2016). Components of effortful control and their relations to children's shyness. *International Journal of Behavioral Development*, 40(6), 544–554. <https://doi.org/10.1177/0165025415597792>
- Elfhag, K., & Morey, L. C. (2008). Personality traits and eating behavior in the obese: Poor self-control in emotional and external eating but personality assets in restrained eating. *Eating Behaviors*, 9(3), 285–293. <https://doi.org/10.1016/j.eatbeh.2007.10.003>
- Emerson, R. W. (2021). Convenience sampling revisited: embracing its limitations through thoughtful study design. *Journal of Visual Impairment & Blindness*, 115(1), 76–77. <https://doi.org/10.1177/0145482X20987707>
- Englert, C. (2021). The strength model of self-control. In *Motivation and self-regulation in sport and exercise* (pp. 102-112). Routledge.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16(1), 143–149. <https://doi.org/10.3758/BF03203267>
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en medición*, 6(1), 27–36.
- Essau, C. A., & de la Torre-Luque, A. (2023). Comorbidity between internalising and externalising disorders among adolescents: Symptom connectivity features and psychosocial outcome. *Child Psychiatry and Human Development*, 54(2), 493–507. <https://doi.org/10.1007/s10578-021-01264-w>

- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Fennig, M. (2023). Adapting Radically Open Dialectical Behavior Therapy (RO-DBT) for adolescents: Preliminary testing of mechanisms of change. [Doctoral dissertation, Washington University in St. Louis]. <https://doi.org/10.7936/4y1r-9a92>
- Fernández-Gómez, E., Martín-Salvador, A., Luque-Vara, T., Sánchez-Ojeda, M. A., Navarro-Prado, S., & Enrique-Mirón, C. (2020). Content validation through expert judgement of an instrument on the nutritional knowledge, beliefs, and habits of pregnant women. *Nutrients*, 12(4), 1136. <https://doi.org/10.3390/nu12041136>
- Finkel, E. J., & Campbell, W. K. (2001). Self-control and accommodation in close relationships: an interdependence analysis. *Journal of personality and social psychology*, 81(2), 263–277. <https://doi.org/10.1037//0022-3514.81.2.263>
- Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In Hancock, G. R., Mueller, R. O. (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 439–492). IAP Information Age Publishing.
- Fino, E., Melogno, S., Iliceto, P., D’Aliesio, S., Pinto, M. A., Candilera, G., & Sabatello, U. (2014). Executive functions, impulsivity, and inhibitory control in adolescents: A structural equation model. *Advances in Cognitive Psychology*, 10(2), 32–38. <https://doi.org/10.5709/acp-0154-5>
- First, M., Gibbon, M., Spritzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *The structured clinical interview for DSM-IV Axis II personality disorders (SCID-II)*. American Psychiatric Press.
- First, M., Williams, J., Karg, R., & Spitzer, R. (2015). *Structured clinical interview for DSM-5, research version*. American Psychiatric Association.
- Fischer, S., Smith, G. T., & Cyders, M. A. (2008). Another look at impulsivity: A meta-analytic review comparing specific dispositions to rash action in their relationship to bulimic symptoms. *Clinical Psychology Review*, 28(8), 1413–1425. <https://doi.org/10.1016/j.cpr.2008.09.001>
- Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates.

- Advances in Methods and Practices in Psychological Science*, 3(4), 484–501.
<https://doi.org/10.1177/2515245920951747>
- Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble. *European Journal of Psychological Assessment*, 33(6), 399–402. <https://doi.org/10.1027/1015-5759/a000460>
- Fonzi, A., Menesini, E. (2005). Coping strategies and resilience' characteristic in adolescence. *Psicologia Clinica dello Sviluppo*, 3, 437- 456.
- Forestier, C., Sarrazin, P., Allenet, B., Gauchet, A., Heuzé, J.-P., & Chalabaev, A. (2018). “Are you in full possession of your capacity?”. A mechanistic self-control approach at trait and state levels to predict different health behaviors. *Personality and Individual Differences*, 134, 214–221.
<https://doi.org/10.1016/j.paid.2018.05.044>
- Fowles, D. C. (1993). Biological variables in psychopathology: A psychobiological perspective. In Adams, H. E., Sutker, P. B. (Eds.), *Comprehensive handbook of psychopathology* (2nd ed., pp. 57–82). Plenum Press.
https://doi.org/10.1007/978-1-4615-3008-4_4
- Freud, S. (1930). *Civilization and its discontents*. Hogarth.
- Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology* (p. 153). Sage Publications Ltd.
- Gailliot, M. T., & Baumeister, R. F. (2007). The physiology of willpower: Linking blood glucose to self-control. *Personality and Social Psychology Review*, 11(4), 303–327. <https://doi.org/10.1177/1088868307303030>
- Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., Plant, E. A., Tice, D. M., Brewer, L. E., & Schmeichel, B. J. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, 92(2), 325–336.
<https://doi.org/10.1037/0022-3514.92.2.325>
- Galupo, M. P., Pulice-Farrow, L., & Ramirez, J. L. (2017). ‘Like a constantly flowing river’: Gender identity flexibility among nonbinary transgender individuals. In J. D. Sinnott (Ed.), *Identity flexibility during adulthood: Perspectives in adult development* (pp. 163–177). Springer International Publishing/Springer Nature. https://doi.org/10.1007/978-3-319-55658-1_10
- General Medical Council. (n.d.). *Annex A: Survey design best practice*. Retrieved July 23, 2023, from https://www.gmc-uk.org/-/media/documents/annex-a-survey-design-best-practice_pdf-79781421.pdf

- Gathergood, J. (2012). Self-control, financial literacy and consumer over-indebtedness. *Journal of Economic Psychology*, 33(3), 590–602. <https://doi.org/10.1016/j.joep.2011.11.006>
- Gerber, E. R., & Wellens, T. R. (1997). Perspectives on pretesting: ‘Cognition’ in the cognitive interview? *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 55(1), 18–39. <https://doi.org/10.1177/075910639705500104>
- Gerbing, D. W., & Hamilton, J. G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. *Structural Equation Modeling*, 3(1), 62–72. <https://doi.org/10.1080/10705519609540030>
- Gibson, C. L., Ward, J. T., Wright, J. P., Beaver, K. M., & Delisi, M. (2010). Where does gender fit in the measurement of self-control? *Criminal Justice and Behavior*, 37(8), 883–903. <https://doi.org/10.1177/0093854810369082>
- Gilbert, K., Barch, D. M., & Luby, J. L. (2020). The Overcontrol in Youth Checklist (OCYC): Behavioral and Neural Validation of a Parent-Report of Child Overcontrol in Early Childhood. *Child Psychiatry & Human Development*, 51(1), 27–38. <https://doi.org/10.1007/s10578-019-00907-3>
- Gilbert, K., Codd III, R. T., Hoyniak, C., Tillman, R., Baudinet, J., Pires, P. P., Hempel, R., Russell, I., & Lynch, T. R. (2023). Processes of change in a randomized clinical trial of radically open dialectical behavior therapy (RO DBT) for adults with treatment-refractory depression. *Journal of Consulting and Clinical Psychology*, 91(2), 71–81. <https://doi.org/10.1037/ccp0000795>
- Gilbert, K., Hall, K., & Codd III, R. T. (2020). Radically Open Dialectical Behavior Therapy: Social signaling, transdiagnostic utility and current evidence. *Psychology Research and Behavior Management*, 13, 19–28. <https://doi.org/10.2147/PRBM.S201848>
- Gilbert, K., Perino, M. T., Myers, M. J., & Sylvester, C. M. (2020). Overcontrol and neural response to errors in pediatric anxiety disorders. *Journal of Anxiety Disorders*, 72, Article 102224. <https://doi.org/10.1016/j.janxdis.2020.102224>
- Gilbert, K., Sudit, E., Fox, N. A., Barch, D. M., & Luby, J. L. (2022). Childhood behavioral inhibition and overcontrol: Relationships with cognitive functioning, error monitoring, anxiety, and obsessive-compulsive symptoms. *Research on Child and Adolescent Psychopathology*, 50(12), 1629–1642. <https://doi.org/10.1007/s10802-022-00953-x>

- Gillebaart, M. (2018). The 'operational' definition of self-control. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01231>
- Gillebaart, M., & De Ridder, D. T. D. (2015). Effortless self-control: A novel perspective on response conflict strategies in trait self-control. *Social and Personality Psychology Compass*, 9(2), 88–99. <https://doi.org/10.1111/spc3.12160>
- Gnanavel, S., Sharma, P., Kaushal, P., & Hussain, S. (2019). Attention deficit hyperactivity disorder and comorbidity: A review of literature. *World Journal of Clinical Cases*, 7(17), 2420–2426. <https://doi.org/10.12998/wjcc.v7.i17.2420>
- Goldberg, L. R. (1990). An alternative 'description of personality': The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59(6), 1216–1229. <https://doi.org/10.1037/0022-3514.59.6.1216>
- Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology*, 48(1), 82–98. <https://doi.org/10.1037/0022-3514.48.1.82>
- Gottfredson, M. R., & Hirschi, T. (1990). *A General Theory of Crime*. Stanford University Press.
- Grant, J. S., & Kinney, M. R. (1992). Using the Delphi technique to examine the content validity of nursing diagnoses. *Nursing Diagnosis*, 3(1), 12–22. <https://doi.org/10.1111/j.1744-618x.1992.tb00193.x>
- Grasmick, H. G., Tittle, C. R., Bursik, R. J., & Arneklev, B. J. (1993). Testing the core empirical implications of Gottfredson and Hirschi's General Theory of Crime. *Journal of Research in Crime and Delinquency*, 30(1), 5–29. <https://doi.org/10.1177/0022427893030001002>
- Gray, J. A. (1991). The neuropsychology of temperament. In J. Strelau & A. Angleitner (Eds.), *Explorations in temperament: International perspectives on theory and measurement* (pp. 105–128). Plenum Press. https://doi.org/10.1007/978-1-4899-0643-4_8
- Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*. Oxford University Press.
- Gray, N. S., Weidacker, K., & Snowden, R. J. (2019). Psychopathy and impulsivity: The relationship of psychopathy to different aspects of UPPS-P impulsivity.

- Psychiatry Research*, 272, 474–482.
<https://doi.org/10.1016/j.psychres.2018.12.155>
- Green, J., & Spikins, P. (2020). Not just a virtue: The evolution of self-control. *Time and Mind*, 13(2), 117–139. <https://doi.org/10.1080/1751696X.2020.1747246>
- Ha, C., Balderas, J. C., Zanarini, M. C., Oldham, J., & Sharp, C. (2014). Psychiatric comorbidity in hospitalized adolescents with borderline personality disorder. *Journal of Clinical Psychiatry*, 75(5), 457–464.
<https://doi.org/10.4088/JCP.13m08696>
- Hagger, M. S. (2010). Sleep, self-regulation, self-control and health. *Stress and Health*, 26(3), 181–185. <https://doi.org/10.1002/smi.1345>
- Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, 136(4), 495–525. <https://doi.org/10.1037/a0019486>
- Hair, J. F., Babin, B. J., Black, W. C., & Anderson, R. E. (2019). *Multivariate Data Analysis*. Cengage.
- Halek, M., Holle, D., & Bartholomeyczik, S. (2017). Development and evaluation of the content validity, practicability and feasibility of the Innovative dementia-oriented Assessment system for challenging behaviour in residents with dementia. *BMC Health Services Research*, 17(1), 554.
<https://doi.org/10.1186/s12913-017-2469-8>
- Hamilton, L., Bacon, L., & Longfellow, E. (2021). Radically Open-Dialectical Behaviour Therapy: A new treatment of people with maladaptive overcontrol who offend. In Winder, B., Blagden, N., Hamilton, L., & Scott, S. (Eds.), *Forensic interventions for therapy and rehabilitation: Case studies and analysis* (pp. 90-107). Routledge.
- Hamilton, L. J. (2021). *When a good thing goes bad: Using personality theory to reconceptualise overcontrolled pathways to offending* [Doctoral dissertation, Nottingham Trent University].
- Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018). A data-driven analysis of workers' earnings on Amazon Mechanical Turk. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14. <https://doi.org/10.1145/3173574.3174023>
- Harrington, D. (2009). *Confirmatory Factor Analysis*. Oxford University Press.
- Hathaway, S. R., & McKinley, J. C. (1967). *The Minnesota Multiphasic Personality Inventory manual*. Psychological Corporation.

- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191–205.
<https://doi.org/10.1177/1094428104263675>
- Heath, N., Midkiff, M. F., Gerhart, J., & Turow, R. G. (2021). Group-based DBT skills training modules are linked to independent and additive improvements in emotion regulation in a heterogeneous outpatient sample. *Psychotherapy Research*, 31(8), 1001–1011. <https://doi.org/10.1080/10503307.2021.1878306>
- Hempel, R. J., Booth, R., Giblin, A., Hamilton, L., Hoch, A., Portner, J., Tomcik, N., Rushbrook, S. C., Simic, M., & Hunt, K. (2018). The implementation of RO DBT in clinical practice. *Behavior Therapist*, 41(3), 161–173.
- Hershorn, M., & Rosenbaum, A. (1991). Over- vs. undercontrolled hostility: Application of the construct to the classification of maritally violent men. *Violence and Victims*, 6(2), 151–158. <https://doi.org/10.1891/0886-6708.6.2.151>
- Heubeck, B. G., Wilkinson, R. B., & Cologon, J. (1998). A second look at Carver and White's (1994) BIS/BAS scales. *Personality and Individual Differences*, 25(4), 785–800. [https://doi.org/10.1016/S0191-8869\(98\)00124-X](https://doi.org/10.1016/S0191-8869(98)00124-X)
- Heym, N., Ferguson, E., & Lawrence, C. (2008). An evaluation of the relationship between Gray's revised RST and Eysenck's PEN: Distinguishing BIS and FFFS in Carver and White's BIS/BAS scales. *Personality and Individual Differences*, 45(8), 709–715. <https://doi.org/10.1016/j.paid.2008.07.013>
- Higgins, G. E. (2007). Examining the original Grasmick scale: A Rasch model approach. *Criminal Justice and Behavior*, 34(2), 157–178.
<https://doi.org/10.1177/0093854806290071>
- Hilton, C. E. (2017). The importance of pretesting questionnaires: A field research example of cognitive pretesting the Exercise referral Quality of Life Scale (ER-QLS). *International Journal of Social Research Methodology*, 20(1), 21–34. <https://doi.org/10.1080/13645579.2015.1091640>
- Hirschi, T., & Gottfredson, M. (1988). Towards a general theory of crime. In W. Buikhuisen, & S. A. Mednick (Eds.), *Explaining criminal behaviour: Interdisciplinary approaches* (pp. 8–26). Brill Sense.
- Hoekstra, R., Vugteveen, J., Warrens, M. J., & Kruijten, P. M. (2019). An empirical analysis of alleged misunderstandings of coefficient alpha. *International*

- Journal of Social Research Methodology*, 22(4), 351–364.
<https://doi.org/10.1080/13645579.2018.1547523>
- Hofmann, W., Luhmann, M., Fisher, R. R., Vohs, K. D., & Baumeister, R. F. (2014). Yes, but are they happy? Effects of trait self-control on affective well-being and life satisfaction. *Journal of Personality*, 82(4), 265–277.
<https://doi.org/10.1111/jopy.12050>
- Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, 1(4), 304–309. <https://doi.org/10.1038/1124>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* (6)2, 65–70.
- Hoppe, C. M., & Singer, R. D. (1976). Overcontrolled hostility, empathy, and egocentric balance in violent and nonviolent psychiatric offenders. *Psychological Reports*, 39(3, Pt 2), 1303–1308.
<https://doi.org/10.2466/pr0.1976.39.3f.1303>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. <https://doi.org/10.1007/BF02289447>
- House, R., Jarvis, N., & Burdsey, D. (2022). Representation matters: Progressing research in plurisexuality and bisexuality in sport. *Journal of Homosexuality*, 69(8), 1301–1321. <https://doi.org/10.1080/00918369.2021.1913916>
- Hoyle, R. H., & Davisson, E. K. (2017). Measurement of self-control by self-report: Considerations and recommendations. In De Ridder, D., Adriaanse, M., & Fujita, K., *Routledge international handbook of self-control in health and well-being* (pp. 74–87). Routledge.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research*, 103(2), 219–230.
<https://doi.org/10.1007/s11205-011-9843-4>
- Hughes, D. J. (2018). Psychometric validity: Establishing the accuracy and appropriateness of psychometric measures. In Irwing, P., Booth, T., & Hughes, D., *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 751–779). Wiley Blackwell. <https://doi.org/10.1002/9781118489772.ch24>

- Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior*, 18(6), 667–683. [https://doi.org/10.1002/\(SICI\)1099-1379\(199711\)18:6<667::AID-JOB874>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1099-1379(199711)18:6<667::AID-JOB874>3.0.CO;2-T)
- Hutton, H. E., Miner, M. H., Blades, J. R., & Langfeldt, V. C. (1992). Ethnic differences on the MMPI Overcontrolled-Hostility Scale. *Journal of Personality Assessment*, 58(2), 260–268. https://doi.org/10.1207/s15327752jpa5802_5
- Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating models of self-regulation. *Annual Review of Psychology*, 72, 319–345. <https://doi.org/10.1146/annurev-psych-061020-105721>
- Isaksson, M., Ghaderi, A., Ramklint, M., & Wolf-Arehult, M. (2021). Radically Open Dialectical Behavior Therapy for anorexia nervosa: A multiple baseline single-case experimental design study across 13 cases. *Journal of Behavior Therapy and Experimental Psychiatry*, 71, Article 101637. <https://doi.org/10.1016/j.jbtep.2021.101637>
- Isaksson, M., Ghaderi, A., Wolf-Arehult, M., Öster, C., & Ramklint, M. (2021). Sharing and connecting with others—Patient experiences of radically open dialectical behavior therapy for anorexia nervosa and overcontrol: A qualitative study. *Journal of Eating Disorders*, 9, Article 29. <https://doi.org/10.1186/s40337-021-00382-z>
- Isaksson, M., Ghaderi, A., Wolf-Arehult, M., & Ramklint, M. (2021). Psychometric properties of the Swedish version of the Ego Resilience scale (ER) and a new shortened version of the Ego Undercontrol Scale (EUC). *Current Psychology*, 40(3), 1498–1506. <https://doi.org/10.1007/s12144-018-0069-3>
- Jasinska, A. J., Yasuda, M., Burant, C. F., Gregor, N., Khatri, S., Sweet, M., & Falk, E. B. (2012). Impulsivity and inhibitory control deficits are associated with unhealthy eating in young adults. *Appetite*, 59(3), 738–747. <https://doi.org/10.1016/j.appet.2012.08.001>
- Johnson, B. N., Ashe, M. L., & Wilson, S. J. (2017). Self-control capacity as a predictor of borderline personality disorder features, problematic drinking, and their co-occurrence. *Journal of Personality Disorders*, 31(3), 289–305. https://doi.org/10.1521/pedi_2016_30_249

- Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, 35(4), 401–415.
<https://doi.org/10.1007/BF02291817>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36.
<https://doi.org/10.1007/BF02291575>
- Kalkbrenner, M. T. (2023). Alpha, omega, and H internal consistency reliability estimates: Reviewing these options and when to use them. *Counseling Outcome Research and Evaluation*, 14(1), 77–88.
<https://doi.org/10.1080/21501378.2021.1940118>
- Kan, I. P., & Drummey, A. B. (2018). Do imposters threaten data quality? An examination of worker misrepresentation and downstream consequences in amazon’s mechanical turk workforce. *Computers in Human Behavior*, 83, 243–253. <https://doi.org/10.1016/j.chb.2018.02.005>
- Kandler, C., & Zapko-Willmes, A. (2017). Theoretical perspectives on the interplay of nature and nurture in personality development. In J. Specht (Ed.), *Personality development across the lifespan* (pp. 101–115). Academic Press.
<https://doi.org/10.1016/B978-0-12-804674-6.00008-9>
- Kane, M. (2009). Validating the interpretations and uses of test scores. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 39–64). IAP Information Age Publishing.
- Kasch, K. L., Rottenberg, J., Arnow, B. A., & Gotlib, I. H. (2002). Behavioral activation and inhibition systems and the severity and course of depression. *Journal of Abnormal Psychology*, 111(4), 589–597.
<https://doi.org/10.1037//0021-843x.111.4.589>
- Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and Amazon’s Mechanical Turk. *Journal of Advertising*, 46(1), 141–155.
<https://doi.org/10.1080/00913367.2016.1269304>
- Kimbrel, N. A., Nelson-Gray, R. O., & Mitchell, J. T. (2007). Reinforcement sensitivity and maternal style as predictors of psychopathology. *Personality and Individual Differences*, 42(6), 1139–1149. <https://doi.org/10.1016/j.paid.2006.06.028>
- Koçar, H., & Koçar, E. Y., (2015). Comparison of different estimation methods for categorical and ordinal data in confirmatory factor analysis. *Journal of Measurement and Evaluation in Education and Psychology*, 6(2), 351–364.

- Koller, I., Levenson, M. R., & Glück, J. (2017). What do you think you are measuring? A mixed-methods procedure for assessing the content validity of test items and theory-based scaling. *Frontiers in Psychology*, 8, Article 126. <https://doi.org/10.3389/fpsyg.2017.00126>
- Kosman, A. (2020). Self-Knowledge and self-control in Plato's Charmides. In F. Leigh (Ed.), *Self-Knowledge in ancient philosophy: The eighth keeling colloquium in ancient philosophy* (pp. 71-86). Oxford University Press. <https://doi.org/10.1093/oso/9780198786061.003.0003>
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477. <https://doi.org/10.1037/abn0000258>
- Kremen, A. M., & Block, J. (1998). The roots of ego-control in young adulthood: Links with parenting in early childhood. *Journal of Personality and Social Psychology*, 75(4), 1062–1075. <https://doi.org/10.1037/0022-3514.75.4.1062>
- Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). The measurement of attitudes. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 21–76). Lawrence Erlbaum Associates Publishers.
- Krueger, R. F., Caspi, A., Moffitt, T. E., White, J., & Stouthamer-Loeber, M. (1996). Delay of gratification, psychopathology, and personality: Is low self-control specific to externalizing problems? *Journal of Personality*, 64(1), 107–129. <https://doi.org/10.1111/j.1467-6494.1996.tb00816.x>
- Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology*, 2, 111–133. <https://doi.org/10.1146/annurev.clinpsy.2.022305.095213>
- Krueger, R. F., & Markon, K. E. (2011). A dimensional-spectrum model of psychopathology: Progress and opportunities. *Archives of General Psychiatry*, 68(1), 10–11. <https://doi.org/10.1001/archgenpsychiatry.2010.188>
- Kulas, J. T., Stachowski, A. A., & Haynes, B. A. (2008). Middle response functioning in Likert-responses to personality items. *Journal of Business and Psychology*, 22(3), 251–259. <https://doi.org/10.1007/s10869-008-9064-2>

- Kyriazos, T., & Poga-Kyriazou, M. (2023). Applied psychometrics: Estimator considerations in commonly encountered conditions in CFA, SEM, and EFA practice. *Psychology, 14*(5), 799-828.
<https://doi.org/10.4236/psych.2023.145043>
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293–323.
[https://doi.org/10.1016/0010-0285\(74\)90015-2](https://doi.org/10.1016/0010-0285(74)90015-2)
- Lahat, A., Hong, M., & Fox, N. A. (2011). Behavioural inhibition: Is it a risk factor for anxiety? *International Review of Psychiatry, 23*(3), 248–257.
<https://doi.org/10.3109/09540261.2011.590468>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Lenzner, T. (2012). Effects of survey question comprehensibility on response quality. *Field Methods, 24*(4), 409–428. <https://doi.org/10.1177/1525822X12448166>
- Leone, L., & Russo, P. M. (2009). Components of the behavioral activation system and functional impulsivity: A test of discriminant hypotheses. *Journal of Research in Personality, 43*(6), 1101–1104.
<https://doi.org/10.1016/j.jrp.2009.08.004>
- Letzring, T. D., Block, J., & Funder, D. C. (2005). Ego-control and ego-resiliency: Generalization of self-report scales based on personality descriptions from acquaintances, clinicians, and the self. *Journal of Research in Personality, 39*(4), 395–422. <https://doi.org/10.1016/j.jrp.2004.06.003>
- Levy, F. (2010). Internalizing versus externalizing comorbidity: Neural circuit hypothesis. *Australian and New Zealand Journal of Psychiatry, 44*(5), 399–409. <https://doi.org/10.3109/00048670903559585>
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*, 55.
- Lilienfeld, S. O. (2003). Comorbidity between and within childhood externalizing and internalizing disorders: Reflections and directions. *Journal of Abnormal Child Psychology, 31*(3), 285–291. <https://doi.org/10.1023/a:1023229529866>

- Lindner, C., Nagy, G., & Retelsdorf, J. (2015). The dimensionality of the Brief Self-Control Scale—An evaluation of unidimensional and multidimensional applications. *Personality and Individual Differences*, 86, 465–473. <https://doi.org/10.1016/j.paid.2015.07.006>
- Lissitz, R. W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60(1), 10–13. <https://doi.org/10.1037/h0076268>
- Little, J. N., & Codd III, R. T. (2020). Radically Open Dialectical Behavior Therapy (RO DBT) in the treatment of perfectionism: A case study. *Journal of Clinical Psychology*, 76(11), 2097–2108. <https://doi.org/10.1002/jclp.23062>
- Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>
- Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, 91(3), 295–327. <https://doi.org/10.1037/0033-295X.91.3.295>
- Loranger, A. W., Janca, A., & Sartorius, N. (1997). *Assessment and Diagnosis of Personality Disorders: The ICD-10 International Personality Disorder Examination (IPDE)*. Cambridge University Press.
- Loxton, N. J., & Dawe, S. (2001). Alcohol abuse and dysfunctional eating in adolescent girls: The influence of individual differences in sensitivity to reward and punishment. *International Journal of Eating Disorders*, 29(4), 455–462. <https://doi.org/10.1002/eat.1042>
- Ludwig, V. U., Stelzel, C., Krutiak, H., Prunkl, C. E., Steimke, R., Paschke, L. M., Kathmann, N., & Walter, H. (2013). Impulsivity, self-control, and hypnotic suggestibility. *Consciousness and Cognition: An International Journal*, 22(2), 637–653. <https://doi.org/10.1016/j.concog.2013.04.001>
- Luoma, J., Codd III, T., & Lynch, T. (2018). *Radically Open Dialectical Behavior Therapy: Shared Features and Differences With ACT, DBT, and CFT*. *Behavior Therapist*, 41(3), 142–153.
- Lynam, D. R., Smith, G. T., Whiteside, S. P., & Cyders, M. A. (2006). The UPPS-P: Assessing five personality pathways to impulsive behavior. *West Lafayette, IN: Purdue University*, 10.
- Lynam, D. R. (2013). Development of a short form of the UPPS-P Impulsive Behavior Scale. *Unpublished technical report*.

- Lynch, T. R. (2018). *Radically Open Dialectical Behavior Therapy: Theory and practice for treating disorders of overcontrol*. New Harbinger Publications.
- Lynch, T. R., Gray, K. L. H., Hempel, R. J., Titley, M., Chen, E. Y., & O'Mahen, H. A. (2013). Radically open-dialectical behavior therapy for adult anorexia nervosa: Feasibility and outcomes from an inpatient program. *BMC Psychiatry*, 13, 1-17. <https://doi.org/10.1186/1471-244X-13-293>
- Lynch, T. R., Hempel, R. J., Whalley, B., Byford, S., Chamba, R., Clarke, P., Clarke, S., Kingdon, D. G., O'Mahen, H., Remington, B., Rushbrook, S. C., Shearer, J., Stanton, M., Swales, M., Watkins, A., & Russell, I. T. (2020). Refractory depression – mechanisms and efficacy of Radically Open Dialectical Behaviour Therapy (RefrAMED): Findings of a randomised trial on benefits and harms. *British Journal of Psychiatry*, 216(4), 204–212. <https://doi.org/10.1192/bjp.2019.53>
- Lynch, T., Hempel, R., Whalley, B., Byford, S., Chamba, R., Clarke, P., Clarke, S., Kingdon, D., O'Mahen, H., Remington, R. and Rushbrook, S., Shearer, J., Stanton, M., Swales, M., Watkins, A., & Russell, I. T. (2018). *Radically Open Dialectical Behaviour Therapy for refractory depression: the RefraMED RCT*. NIHR Journals Library. <https://doi.org/10.3310/eme05070>
- Lynch, T. R., Whalley, B., Hempel, R. J., Byford, S., Clarke, P., Clarke, S., Kingdon, D., O'Mahen, H., Russell, I. T., Shearer, J., Stanton, M., Swales, M., Watkins, A., & Remington, B. (2015). Refractory depression: Mechanisms and evaluation of Radically Open Dialectical Behaviour Therapy (RO-DBT) [REFRAMED]: protocol for randomised trial. *BMJ Open*, 5(7), Article e008857. <https://doi.org/10.1136/bmjopen-2015-008857>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385. <https://doi.org/10.1097/00006199-198611000-00017>
- MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., Aureli, F., Baker, J. M., Bania, A. E., Barnard, A. M., Boogert, N. J., Brannon, E. M., Bray, E. E., Bray, J., Brent, L. J. N., Burkart, J. M., Call, J., Cantlon, J. F., Cheke, L. G., . . . Zhao, Y. (2014). The evolution of self-control. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 111(20), 2140–2148. <https://doi.org/10.1073/pnas.1323533111>

- Maghbouleh, N., Schachter, A., & Flores, R. D. (2022). Middle Eastern and North African Americans may not be perceived, nor perceive themselves, to be White. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 119(7), 1–9. <https://doi.org/10.1073/pnas.2117940119>
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7(6), 537–542. <https://doi.org/10.1177/1745691612460688>
- Manapat, P. D., Edwards, M. C., MacKinnon, D. P., Poldrack, R. A., & Marsch, L. A. (2021). A psychometric analysis of the Brief Self-Control Scale. *Assessment*, 28(2), 395–412. <https://doi.org/10.1177/1073191119890021>
- Mao, T., Pan, W., Zhu, Y., Yang, J., Dong, Q., & Zhou, G. (2018). Self-control mediates the relationship between personality trait and impulsivity. *Personality and Individual Differences*, 129, 70–75. <https://doi.org/10.1016/j.paid.2018.03.013>
- Marcus, B. (2004). Self-control in the General Theory of Crime: Theoretical implications of a measurement problem. *Theoretical Criminology*, 8(1), 33–55. <https://doi.org/10.1177/1362480604039740>
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, 16(3), 439–476. <https://doi.org/10.1080/10705510903008220>
- Maruish, M. E. (2014). The clinical interview. In Archer R. P., & Smith S. R., *Personality Assessment* (2nd ed., pp. 37–88). Routledge.
- Matta, A. da, Gonçalves, F. L., & Bizarro, L. (2012). Delay discounting: Concepts and measures. *Psychology & Neuroscience*, 5, 135–146. <https://doi.org/10.3922/j.psns.2012.2.03>
- Mauss, I. B., Shallcross, A. J., Troy, A. S., John, O. P., Ferrer, E., Wilhelm, F. H., & Gross, J. J. (2011). Don't hide your happiness! Positive emotion dissociation, social connectedness, and psychological functioning. *Journal of Personality and Social Psychology*, 100(4), 738–748. <https://doi.org/10.1037/a0022410>
- McConaughy, S. H., & Skiba, R. J. (1993). Comorbidity of externalizing and internalizing problems. *School Psychology Review*, 22(3), 421–436.

- McCrae, R. R., & John, O. P. (1992). An introduction to the Five-Factor Model and its applications. *Journal of Personality*, 60(2), 175–215.
<https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, 15(1), 28–50.
<https://doi.org/10.1177/1088868310366253>
- McCreary, C., & Padilla, E. (1977). MMPI differences among Black, Mexican-American, and White male offenders. *Journal of Clinical Psychology*, 33(1), 171–177. [https://doi.org/10.1002/1097-4679\(197701\)33:1+<171::AID-JCLP2270330138>3.0.CO;2-Q](https://doi.org/10.1002/1097-4679(197701)33:1+<171::AID-JCLP2270330138>3.0.CO;2-Q)
- McDonald, R. P. (2013). *Test theory: A Unified Treatment*. Psychology Press.
- McGurk, B. J., & McGurk, R. E. (1979). Personality types among prisoners and prison officers: An Investigation of Megargee's Theory of Control. *The British Journal of Criminology*, 19(1), 31–49.
- McHugh, C., & Balaratnasingam, S. (2018). Impulsivity in personality disorders: Current views and future directions. *Current Opinion in Psychiatry*, 31(1), 63.
<https://doi.org/10.1097/YCO.0000000000000383>
- McNaughton, N., & Corr, P. J. (2008). The neuropsychology of fear and anxiety: A foundation for Reinforcement Sensitivity Theory. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality* (pp. 44–94). Cambridge University Press. <https://doi.org/10.1017/CBO9780511819384.003>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, 28(1), 61–88.
<https://doi.org/10.1037/met0000425>
- McNeish, D., & Wolf, M. G. (2024). Direct discrepancy dynamic fit index cutoffs for arbitrary covariance structure models. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–28.
<https://doi.org/10.1080/10705511.2024.2308005>
- Megargee, E. I. (1966). Undercontrolled and overcontrolled personality types in extreme antisocial aggression. *Psychological Monographs: General and Applied*, 80(3), 1–29. <https://doi.org/10.1037/h0093894>

- Megargee, E. I., Cook, P. E., & Mendelsohn, G. A. (1967). Development and validation of an MMPI scale of assaultiveness in overcontrolled individuals. *Journal of Abnormal Psychology, 72*(6), 519–528.
<https://doi.org/10.1037/h0025242>
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56*(2), 128–165. <https://doi.org/10.1037/0003-066X.56.2.128>
- Meyer, M. J. (1987). Stoics, rights, and autonomy. *American Philosophical Quarterly, 24*(3), 267–271.
- Milioni, M., Alessandri, G., Eisenberg, N., Vecchione, M., & Caprara, G. V. (2015). The predictive role of ego-resiliency on behavioural problems. *European Journal of Developmental Psychology, 12*(2), 220–233.
<https://doi.org/10.1080/17405629.2014.998194>
- Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology, 54*(4), 687–696.
<https://doi.org/10.1037/0022-3514.54.4.687>
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, pp. 2693–2698.
<https://doi.org/10.1073/pnas.1010076108>
- Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica, 30*(1), Article 3.
- Morgan, D. L. (1996). Focus groups. *Annual Review of Sociology, 22*, 129–152.
- Möttus, R., Realo, A., Allik, J., Ausmees, L., Henry, S., McCrae, R. R., & Vainik, U. (2024). Most people's life satisfaction matches their personality traits: True correlations in multi-trait, multi-rater, multi-sample data. *Journal of Personality and Social Psychology, 126*(4), 676–693.
<https://doi.org/10.1037/pspp0000501>

- Muraven, M. (2010). Building self-control strength: Practicing self-control leads to improved self-control performance. *Journal of Experimental Social Psychology*, 46(2), 465–468. <https://doi.org/10.1016/j.jesp.2009.12.011>
- Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126(2), 247–259. <https://doi.org/10.1037/0033-2909.126.2.247>
- Muraven, M., Baumeister, R. F., & Tice, D. M. (1999). Longitudinal improvement of self-regulation through practice: building self-control strength through repeated exercise. *Journal of Social Psychology*, 139(4), 446–457. <https://doi.org/10.1080/00224549909598404>
- Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, 74(3), 774–789. <https://doi.org/10.1037/0022-3514.74.3.774>
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications*. Pearson/Prentice-Hall.
- Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *Journal of General Psychology*, 142(2), 71–89. <https://doi.org/10.1080/00221309.2014.994590>
- Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants. *PLoS ONE*, 11(6), Article e0157732. <https://doi.org/10.1371/journal.pone.0157732>
- Neuberg, S. L., & Newsom, J. T. (1993). Personal need for structure: Individual differences in the desire for simpler structure. *Journal of Personality and Social Psychology*, 65(1), 113–131. <https://doi.org/10.1037/0022-3514.65.1.113>
- Newman, I., Lim, J., & Pineda, F. (2013). Content validity using a mixed methods approach: Its application and development through the use of a table of specifications methodology. *Journal of Mixed Methods Research*, 7(3), 243–260. <https://doi.org/10.1177/1558689813476922>
- Nigg, J. T. (2017). Annual research review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 58(4), 361–383. <https://doi.org/10.1111/jcpp.12675>

- Nunnally, J. C. (1978). An overview of psychological measurement. In B. B. Wolman (Ed.), *Clinical diagnosis of mental disorders: A handbook* (pp. 97–146). Springer. https://doi.org/10.1007/978-1-4684-2490-4_4
- Ong, A. D., & Bergeman, C. S. (2004). The complexity of emotions in later life. *Journals of Gerontology: Psychological Sciences and Social Sciences*, 59(3), 117–122. <https://doi.org/10.1093/geronb/59.3.P117>
- Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, 5(3), 343–355. <https://doi.org/10.1037/1082-989X.5.3.343>
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*. University of Illinois Press.
- Oshio, A., Taku, K., Hirano, M., & Saeed, G. (2018). Resilience and Big Five personality traits: A meta-analysis. *Personality and Individual Differences*, 127, 54–60. <https://doi.org/10.1016/j.paid.2018.01.048>
- Padilla, M. (2019). A Primer on Reliability via Coefficient Alpha and Omega. *Archives of Psychology*, 3(8), 1-15.
- Padilla, M. A., & Divers, J. (2016). A comparison of composite reliability estimators: Coefficient omega confidence intervals in the current literature. *Educational and Psychological Measurement*, 76(3), 436–453. <https://doi.org/10.1177/0013164415593776>
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 224–239). The Guilford Press.
- Pearson, C. M., Mason, T. B., Cao, L., Goldschmidt, A. B., Lavender, J. M., Crosby, R. D., Crow, S. J., Engel, S. G., Wonderlich, S. A., & Peterson, C. B. (2018). A test of a state-based, self-control theory of binge eating in adults with obesity. *Eating Disorders: Journal of Treatment & Prevention*, 26(1), 26–38. <https://doi.org/10.1080/10640266.2018.1418358>
- Pechorro, P., DeLisi, M., Pacheco, C., Abrunhosa Gonçalves, R., Maroco, J., & Quintas, J. (2023). Examination of Grasmick et al.'s Low Self-Control Scale and of a short version with cross-gender measurement invariance. *Crime &*

- Delinquency*, 69(13/14), 2741–2764.
<https://doi.org/10.1177/00111287211073674>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163.
<https://doi.org/10.1016/j.jesp.2017.01.006>
- Peer, E., Rothschild, D., & Gordon, A. (2023). *Platform over procedure: Online platforms that pre-vet participants yield higher data quality without sacrificing diversity*. PsyArXiv. <https://doi.org/10.31234/osf.io/buzwn>
- Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 54(4), 1643–1662. <https://doi.org/10.3758/s13428-021-01694-3>
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, 21(2), 381–391. <https://doi.org/10.1086/209405>
- Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology*, 98(1), 194–198.
<https://doi.org/10.1037/a0030767>
- Philippe, F. L., Laventure, S., Beaulieu-Pelletier, G., Lecours, S., & Lekes, N. (2011). Ego-resiliency as a mediator between childhood trauma and psychological symptoms. *Journal of Social and Clinical Psychology*, 30(6), 583–598. <https://doi.org/10.1521/jscp.2011.30.6.583>
- Pilcher, J. J., Morris, D. M., & Erikson, D. N. (2023). Self-control measurement methodologies: An integrative approach. *Psychological Reports*, 126(3), 1108–1129. <https://doi.org/10.1177/00332941211067969>
- Piquero, A. R., MacIntosh, R., & Hickman, M. (2000). Does self-control affect survey response? Applying exploratory, confirmatory, and item response theory analysis to Grasmick et al.'s self-control scale. *Criminology*, 38(3), 897–930. <https://doi.org/10.1111/j.1745-9125.2000.tb00910.x>
- Plomin, R. (1994). *Genetics and experience: The interplay between nature and nurture*. Sage Publications.
- Polit, D. F. (2014). Getting serious about test–retest reliability: A critique of retest research and some recommendations. *International Journal of Quality of Life Aspects of Treatment, Care, & Rehabilitation*, 23(6), 1713–1720.
<https://doi.org/10.1007/s11136-014-0632-9>

- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>
- Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage: A polyvagal theory. *Psychophysiology*, 32(4), 301–318. <https://doi.org/10.1111/j.1469-8986.1995.tb01213.x>
- Porges, S. W. (2001). The polyvagal theory: Phylogenetic substrates of a social nervous system. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 42(2), 123–146. [https://doi.org/10.1016/s0167-8760\(01\)00162-3](https://doi.org/10.1016/s0167-8760(01)00162-3)
- Porges, S. W. (2003). The polyvagal theory: Phylogenetic contributions to social behavior. *Physiology & Behavior*, 79(3), 503–513. [https://doi.org/10.1016/S0031-9384\(03\)00156-2](https://doi.org/10.1016/S0031-9384(03)00156-2)
- Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, 74(2), 116–143. <https://doi.org/10.1016/j.biopsycho.2006.06.009>
- Porges, S. W. (2009). The polyvagal theory: New insights into adaptive reactions of the autonomic nervous system. *Cleveland Clinic Journal of Medicine*, 76(2), 86–90. <https://doi.org/10.3949/ccjm.76.s2.17>
- Post, R. M. (2007). Kindling and sensitization as models for affective episode recurrence, cyclicity, and tolerance phenomena. *Neuroscience and Biobehavioral Reviews*, 31(6), 858–873. <https://doi.org/10.1016/j.neubiorev.2007.04.003>
- Powers, W. T. (1973). Feedback: Beyond behaviorism. *Science*, 179(4071), 351–356. <https://doi.org/10.1126/science.179.4071.351>
- Poythress, N. G., Skeem, J. L., Weir, J., Lilienfeld, S. O., Douglas, K. S., Edens, J. F., & Kennealy, P. J. (2008). Psychometric properties of Carver and White's (1994) BIS/BAS scales in a large sample of offenders. *Personality and Individual Differences*, 45(8), 732–737. <https://doi.org/10.1016/j.paid.2008.07.021>
- Pratt, T. C., & Cullen, F. T. (2000). The empirical status of Gottfredson and Hirschi's General Theory of Crime: A Meta-Analysis. *Criminology*, 38(3), 931–964. <https://doi.org/10.1111/j.1745-9125.2000.tb00911.x>
- Pulkkinen, L. (1996). Female and male personality styles: A typological and developmental analysis. *Journal of Personality and Social Psychology*, 70(6), 1288–1306. <https://doi.org/10.1037/0022-3514.70.6.1288>

- Quadri, N., Wild, D., Skerrett, B., Muehlhausen, W., & O'Donohoe, P. (2013). A literature review of the variance in interval length between administrations for assessment of test-retest reliability and equivalence of pro measures. *Value in Health, 16*(3), A40–A41. <https://doi.org/10.1016/j.jval.2013.03.230>
- Quilty, L. C., & Oakman, J. M. (2004). The assessment of behavioural activation--the relationship between impulsivity and behavioural activation. *Personality and Individual Differences, 37*(2), 429–442. <https://doi.org/10.1016/j.paid.2003.09.014>
- Rahman, F. (2018). Self-control in Islām and its psychological aspect. *Journal of Religious Studies, 2*(3), 16-36. <https://doi.org/10.33195/uochjrs-v11iIII952018>
- Rennie, K. M. (1997). *Exploratory and confirmatory rotation strategies in exploratory factor analysis* [ERIC Document Reproduction Service No. ED406446].
- Reynolds, B., Ortengren, A., Richards, J. B., & De Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences, 40*(2), 305–315. <https://doi.org/10.1016/j.paid.2005.03.024>
- Reynolds, N., Diamantopoulos, A., & Schlegelmilch, B. B. (1993). Pretesting in questionnaire design: A review of the literature and suggestions for further research. *Journal of the Market Research Society, 35*(2), 171–182. <https://doi.org/10.1177/147078539303500202>
- Righetti, F., & Finkenauer, C. (2011). If you are able to control yourself, I will trust you: The role of perceived self-control in interpersonal trust. *Journal of Personality and Social Psychology, 100*(5), 874–886. <https://doi.org/10.1037/a0021827>
- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2*(4), 313–345. <https://doi.org/10.1111/j.1745-6916.2007.00047.x>
- Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin, 143*(2), 117–141. <https://doi.org/10.1037/bul0000088>

- Robins, R. W., John, O. P., Caspi, A., Moffitt, T. E., & Stouthamer-Loeber, M. (1996). Resilient, overcontrolled, and undercontrolled boys: Three replicable personality types. *Journal of Personality and Social Psychology*, 70(1), 157–171. <https://doi.org/10.1037/0022-3514.70.1.157>
- Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Journal of the International Society for Pharmacoeconomics and Outcomes Research*, 14(8), 1101–1108. <https://doi.org/10.1016/j.jval.2011.06.003>
- Rounding, K., Lee, A., Jacobson, J. A., & Ji, L. J. (2012). Religion replenishes self-control. *Psychological science*, 23(6), 635–642. <https://doi.org/10.1177/0956797611431987>
- Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, 25(1), 6–14. <https://doi.org/10.1177/1094428120968614>
- Salazar, M. S. (2015). The dilemma of combining positive and negative items in scales. *Psicothema*, 27(2), 192–199.
- Savvidou, L. G., Fagundo, A. B., Fernández-Aranda, F., Granero, R., Claes, L., Mallorquí-Baqué, N., Verdejo-García, A., Steiger, H., Israel, M., Moragas, L., del Pino-Gutiérrez, A., Aymamí, N., Gómez-Peña, M., Agüera, Z., Tolosa-Sola, I., La Verde, M., Aguglia, E., Menchón, J. M., & Jiménez-Murcia, S. (2017). Is gambling disorder associated with impulsivity traits measured by the UPPS-P and is this association moderated by sex and age? *Comprehensive Psychiatry*, 72, 106–113. <https://doi.org/10.1016/j.comppsy.2016.10.005>
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353. <https://doi.org/10.1037/1040-3590.8.4.350>
- Sellars, J. (2006). An ethics of the event: Deleuze's stoicism. *Angelaki*, 11(3), 157–171. <https://doi.org/10.1080/09697250601048622>
- Serrano-Ibáñez, E. R., Ramírez-Maestre, C., López-Martínez, A. E., Esteve, R., Ruiz-Párraga, G. T., & Jensen, M. P. (2018). Behavioral inhibition and activation systems, and emotional regulation in individuals with chronic musculoskeletal pain. *Frontiers in Psychiatry*, 9, Article 394. <https://doi.org/10.3389/fpsy.2018.00394>
- Shamon, H., & Berning, C. (2019). Attention check items and instructions in online surveys with incentivized and non-incentivized samples: Boon or bane for

- data quality? *Survey Research Methods* 14(1), 55-77.
<https://doi.org/10.18148/srm/2020.v14i1.7374>
- Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using mechanical turk to study clinical populations. *Clinical Psychological Science*, 1(2), 213–220.
<https://doi.org/10.1177/2167702612469015>
- Sharp, K. L., Williams, A. J., Rhyner, K. T., & Ilardi, S. S. (2013). The clinical interview. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 2. Testing and assessment in clinical and counseling psychology* (pp. 103–117). American Psychological Association. <https://doi.org/10.1037/14048-007>
- Shiyanbola, O. O., Bolt, D., Tarfa, A., Brown, C., & Ward, E. (2019). A content validity and cognitive interview process to evaluate an Illness Perception Questionnaire for African Americans with type 2 diabetes. *BMC Research Notes*, 12(1), Article 308. <https://doi.org/10.1186/s13104-019-4342-9>
- Silverstein, A. B. (1987). Note on the parallel analysis criterion for determining the number of common factors or principal components. *Psychological Reports*, 61(2), 351–354. <https://doi.org/10.2466/pr0.1987.61.2.351>
- Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*, 31(4), 557–566.
<https://doi.org/10.1037/pas0000648>
- Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher*, 36(8), 477–481. <https://doi.org/10.3102/0013189X07311609>
- Smith, J. L., Mattick, R. P., Jamadar, S. D., & Iredale, J. M. (2014). Deficits in behavioural inhibition in substance abuse and addiction: A meta-analysis. *Drug and Alcohol Dependence*, 145, 1–33.
<https://doi.org/10.1016/j.drugalcdep.2014.08.009>
- Smith, S. M., Roster, C. A., Golden, L. L., & Albaum, G. S. (2016). A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research*, 69(8), 3139–3148. <https://doi.org/10.1016/j.jbusres.2015.12.002>
- Smits, D. J. M., & Kuppens, P. (2005). The relations between anger, coping with anger, and aggression, and the BIS/BAS system. *Personality and Individual Differences*, 39(4), 783–793. <https://doi.org/10.1016/j.paid.2005.02.023>

- Sonnby-Borgström, M. (2002). Automatic mimicry reactions as related to differences in emotional empathy. *Scandinavian Journal of Psychology*, 43(5), 433–443. <https://doi.org/10.1111/1467-9450.00312>
- Sonnby-Borgström, M., Jönsson, P., & Svensson, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *Journal of Nonverbal Behavior*, 27(1), 3–23. <https://doi.org/10.1023/A:1023608506243>
- Spearman, C. (1910). Correlation calculated from faulty data. *British journal of psychology*, 3(3), 271-295.
- Spiegel, J. S. (2020). Cultivating self-control: Foundations and methods in the Christian theological tradition. *Journal of Spiritual Formation and Soul Care*, 13(2), 193-210. <https://doi.org/10.1177/193979092091888>
- Sportel, B. E., Nauta, M. H., de Hullu, E., & de Jong, P. J. (2013). Predicting internalizing symptoms over a two-year period by BIS, FFFS and attentional control. *Personality and Individual Differences*, 54(2), 236-240. <https://doi.org/10.1016/j.paid.2012.08.043>
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, 91(6), 1292–1306. <https://doi.org/10.1037/0021-9010.91.6.1292>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662. <https://doi.org/10.1037/h0054651>
- Strus, W., Cybis, N., Cieciuch, J., & Rowiński, T. (2021). Theoretical framework for the RUNO personality typology based on the Circumplex of Personality Metatraits. *Polish Psychological Bulletin*, 52(3), 197–210. <https://doi.org/10.24425/ppb.2021.137885>
- Swanson, J., Valiente, C., Lemery-Chalfant, K., & Caitlin O'Brien, T. (2011). Predicting early adolescents' academic achievement, social competence, and physical health from parenting, ego resilience, and engagement coping. *Journal of Early Adolescence*, 31(4), 548–576. <https://doi.org/10.1177/0272431610366249>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>

- Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72(2), 271–322.
<https://doi.org/10.1111/j.0022-3506.2004.00263.x>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.
<https://doi.org/10.5116/ijme.4dfb.8dfd>
- Teitcher, J. E. F., Bockting, W. O., Bauermeister, J. A., Hoefer, C. J., Miner, M. H., & Klitzman, R. L. (2015). Detecting, preventing, and responding to 'fraudsters' in internet research: Ethics and tradeoffs. *Journal of Law, Medicine & Ethics*, 43(1), 116–133. <https://doi.org/10.1111/jlme.12200>
- Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, 61(3), 201–216.
[https://doi.org/10.1016/S0165-0327\(00\)00338-4](https://doi.org/10.1016/S0165-0327(00)00338-4)
- Thayer, J. F., & Lane, R. D. (2009). Claude Bernard and the heart-brain connection: Further elaboration of a model of neurovisceral integration. *Neuroscience and Biobehavioral Reviews*, 33(2), 81–88.
<https://doi.org/10.1016/j.neubiorev.2008.08.004>
- Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, 56(2), 197–208.
<https://doi.org/10.1177/0013164496056002001>
- Thürmer, J. L., Scheier, M. F., & Carver, C. S. (2020). On the mechanics of goal striving: Experimental evidence of coasting and shifting. *Motivation Science*, 6(3), 266–274. <https://doi.org/10.1037/mot0000157>
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286. <https://doi.org/10.1037/h0070288>
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529–554. <https://doi.org/10.1086/214483>
- Thurstone, L. L. (1929). Theory of attitude measurement. *Psychological Review*, 36(3), 222–241. <https://doi.org/10.1037/h0070922>
- Tran, J., Teese, R., & Gill, P. R. (2018). UPPS-P facets of impulsivity and alcohol use patterns in college and noncollege emerging adults. *The American Journal of Drug and Alcohol Abuse*, 44(6), 695–704.
<https://doi.org/10.1080/00952990.2018.1503280>

- Truscott, D. (1990). Assessment of overcontrolled hostility in adolescence. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2(2), 145–148. <https://doi.org/10.1037/1040-3590.2.2.145>
- Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement*, 58(4), 541–568. <https://doi.org/10.1177/0013164498058004001>
- Tyler, J. M., & Burns, K. C. (2008). After depletion: The replenishment of the self's regulatory resources. *Self and Identity*, 7(3), 305–321. <https://doi.org/10.1080/15298860701799997>
- Tyndall, I., Waldeck, D., Pancani, L., Whelan, R., Roche, B., & Dawson, D. L. (2019). The Acceptance and Action Questionnaire-II (AAQ-II) as a measure of experiential avoidance: Concerns over discriminant validity. *Journal of Contextual Behavioral Science*, 12, 278–284. <https://doi.org/10.1016/j.jcbs.2018.09.005>
- Um, M., Hershberger, A. R., Whitt, Z. T., & Cyders, M. A. (2018). Recommendations for applying a multi-dimensional model of impulsive personality to diagnosis and treatment. *Borderline Personality Disorder and Emotion Dysregulation*, 5, Article 6. <https://doi.org/10.1186/s40479-018-0084-x>
- Vanderbleek, E., & Gilbert, K. (2018). Too much versus too little control: The etiology, conceptualization, and treatment implications of overcontrol and undercontrol. *Behavior Therapist*, 41(3), 125-131.
- Vaughn, M. G., DeLisi, M., Beaver, K. M., Wright, J. P., & Howard, M. O. (2007). Toward a psychopathology of self-control theory: The importance of narcissistic traits. *Behavioral Sciences & the Law*, 25(6), 803–821. <https://doi.org/10.1002/bsl.789>
- Vazsonyi, A. T., & Belliston, L. M. (2007). The family → Low self-control → Deviance: A cross-cultural and cross-national test of self-control theory. *Criminal Justice and Behavior*, 34(4), 505–530. <https://doi.org/10.1177/0093854806292299>
- Vazsonyi, A. T., Mikuška, J., & Kelley, E. L. (2017). It's time: A meta-analysis on the self-control-deviance link. *Journal of Criminal Justice*, 48, 48–63. <https://doi.org/10.1016/j.jcrimjus.2016.10.001>

- Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In *Problems and solutions in human assessment* (pp. 41–71). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-4397-8_3
- Venables, N. C., Foell, J., Yancey, J. R., Beaver, K. M., Iacono, W. G., & Patrick, C. J. (2018). Integrating criminological and mental health perspectives on low self-control: A multi-domain analysis. *Journal of Criminal Justice*, 56, 2–10. <https://doi.org/10.1016/j.jcrimjus.2017.10.004>
- Verona, E., Sprague, J., & Sadeh, N. (2012). Inhibitory control and negative emotional processing in psychopathy and antisocial personality disorder. *Journal of Abnormal Psychology*, 121(2), 498–510. <https://doi.org/10.1037/a0025308>
- Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Anales de Psicología*, 33(3), 755–782. <https://doi.org/10.6018/analesps.33.3.268401>
- Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., & Schroeder, M. (2011). Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing*, 3(1), 69-87.
- Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology*, 48(4), 865–885. <https://doi.org/10.1111/j.1744-6570.1995.tb01784.x>
- Waaktaar, T., & Torgersen, S. (2010). How resilient are resilience scales? The Big Five scales outperform resilience scales in predicting adjustment in adolescents. *Scandinavian Journal of Psychology*, 51(2), 157–163. <https://doi.org/10.1111/j.1467-9450.2009.00757.x>
- Ward, J. T., Ray, J. V., & Fox, K. A. (2018). Exploring differences in self-control across sex, race, age, education, and language: Considering a bifactor MIMIC model. *Journal of Criminal Justice*, 56, 29–42. <https://doi.org/10.1016/j.jcrimjus.2017.09.006>
- Wennerhold, L., & Frieze, M. (2020). Why self-report measures of self-control and inhibition tasks do not substantially correlate. *Collabra: Psychology*, 6(1), Article 9. <https://doi.org/10.1525/collabra.276>

- Wennerhold, L., & Frieze, M. (2023). Challenges in the conceptualization of trait self-control as a psychological construct. *Social and Personality Psychology Compass*, 17(3), Article e12726. <https://doi.org/10.1111/spc3.12726>
- Westen, D., DeFife, J. A., Bradley, B., & Hilsenroth, M. J. (2010). Prototype personality diagnosis in clinical practice: A viable alternative for DSM-V and ICD-11. *Professional Psychology, Research and Practice*, 41(6), 482–487. <https://doi.org/10.1037/a0021555>
- Whiteside, S. P., & Lynam, D. R. (2001). The five Factor Model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, 30(4), 669–689. [https://doi.org/10.1016/S0191-8869\(00\)00064-7](https://doi.org/10.1016/S0191-8869(00)00064-7)
- Widiger, T. A., & Clark, L. A. (2000). Toward DSM-V and the classification of psychopathology. *Psychological Bulletin*, 126(6), 946–963. <https://doi.org/10.1037/0033-2909.126.6.946>
- Williams, B., Onsmann, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 8, 1–13. <https://doi.org/10.33151/ajp.8.3.93>
- Willis, G. B. (2004). *Cognitive Interviewing: A tool for improving questionnaire design*. Sage Publications.
- Willis, G. B., & Artino, A. R. (2013). What do our respondents think we're asking? Using cognitive interviewing to improve medical education surveys. *Journal of Graduate Medical Education*, 5(3), 353–356. <https://doi.org/10.4300/JGME-D-13-00154.1>
- Willner, C. J., Gatzke-Kopp, L. M., & Bray, B. C. (2016). The dynamics of internalizing and externalizing comorbidity across the early school years. *Development and Psychopathology*, 28(4pt1), 1033–1052. <https://doi.org/10.1017/S0954579416000687>
- Wolgast, M. (2014). What does the Acceptance and Action Questionnaire (AAQ-II) really measure? *Behavior Therapy*, 45(6), 831–839. <https://doi.org/10.1016/j.beth.2014.07.002>
- Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806–838. <https://doi.org/10.1177/0011000006288127>
- Xu, Y., Pace, S., Kim, J., Iachini, A., King, L. B., Harrison, T., DeHart, D., Levkoff, S. E., Browne, T. A., Lewis, A. A., Kunz, G. M., Reitmeier, M., Utter, R. K.,

- & Simone, M. (2022). Threats to online surveys: Recognizing, detecting, and preventing survey bots. *Social Work Research*, 46(4), 343–350.
<https://doi.org/10.1093/swr/svac023>
- Yang, Y., & Green, S. B. (2010). A note on structural equation modeling estimates of reliability. *Structural Equation Modeling*, 17(1), 66–81.
<https://doi.org/10.1080/10705510903438963>
- Yang, Y., & Xia, Y. (2015). On the number of factors to retain in exploratory factor analysis for ordered categorical data. *Behavior Research Methods*, 47(3), 756–772. <https://doi.org/10.3758/s13428-014-0499-2>
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, 6, 21–29.
<https://doi.org/10.56801/10.56801/v6.i.279>
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99(3), 432–442.
<https://doi.org/10.1037/0033-2909.99.3.432>

Appendices

Appendix A:

A pilot focus group schedule for the item generation study.

Self-control and its components

1. Firstly, I'd like to know what you think about when you hear the term 'self-control'. What are the first things that come to mind?
2. And considering the views we have just heard, would you say you have a lot of self-control, or little? [...] Would you like to elaborate on why, or give us examples?
3. **Inhibition** refers to the idea that people who have good self-control can resist temptations, impulses, and behaviours that could stop them from achieving their goals – such as eating a slice of cake when on a diet. How easy would you say resisting temptations and impulses is for you?
4. Thanks. Now, let's think about **initiation**. Initiation refers to the idea that people who have a lot of self-control can easily initiate behaviours that help them achieve their goals, even if they may be unpleasant – such as eating healthier when trying to lose weight or exercising more often. How easy would you say this is for you?
5. Thank you. Now, the last component is **effort**. I asked you before how easily you can inhibit and initiate behaviours. In other words, I wanted to know how much effort it takes to exercise self-control for you. There are some disagreements between researchers of whether self-control always requires effort, but it has been suggested that for those with good self-control, inhibiting and initiating behaviours often requires less, or even no effort in comparison with those who have less self-control. Any reflections on this?

Receptivity and openness

1. **How open are you to new activities and experiences? Think of a time when you had a chance of trying out something new or experiencing something you've never experienced before. If comfortable, describe the situation to others. How did it make you feel?**
Scenario: Your friend invites you to go for a skiing holiday with them. You have enough leave at work, been saving for a while, and a holiday seems like a good idea. However, you have never been skiing before, and you know there is a high risk of injury. How does this make you feel?
2. **Think of a time when you were faced with uncertainty. How did you manage this situation? How did the uncertainty make you feel?**
Scenario: You did not have time to study for an important exam at university but had to take it anyway. You are not sure how you did, and it is important that you pass. You will not find out the results for another month. How does the uncertainty of this situation make you feel?

3. **How comfortable are you with taking risks? Think of a time when you were faced with an unplanned risk. If comfortable, describe the situation to others. How did you manage this situation? How did it make you feel?**
Scenario: At work, you have just been asked to step in for a sick colleague and deliver a presentation to a new potential customer. You are good at your job, and you know how to sell a product, but you did not have time to prepare beforehand. There would be no consequences if you did not agree to deliver the presentation, but if you do well, there is a good chance you will be rewarded for your achievement. What do you do? Why?
4. **Think of a situation when you received critical feedback from someone. How did receiving critical feedback make you feel? How did you react?**
Scenario: You are working on a project as part of a team. One day, the team leader calls you into their office and you receive some critical feedback on your work. You did not expect this. How does it make you feel?
5. **Do you get excited easily? Think of a time when something good happened or was about to happen. How did it make you feel?**
Scenario: You just found out your favourite artist will soon perform in your city. How do you react? What are your thoughts?
6. **Do you get anxious easily? Think of a time when you felt that something bad or unpleasant might happen. How did it make you feel? How did you manage the situation?**
Scenario: You are on a plane, and the seatbelt button lights up. You put on your seatbelt and start feeling slight turbulence. What's your first thought?

Flexible control

1. **Would you describe yourself as a perfectionist? Think of a time when you were not quite happy with the work you produced. How did that make you feel? What did you do?**
Scenario: You must send a short e-mail to somebody. You try writing it, but you don't think that you are using the right words. You feel like you could have phrased it better, but you just can't decide on how. What do you do?
2. **How sensitive are you to rules and rule-breaking? Think of a time you have or could have broken a rule. How does rule-breaking make you feel?**
Scenario: On your way to the shop, and you want to take a shortcut through a park. On the gate, you read that the park closes to the public at 5pm. It is now 7pm, but the gate is still open. Do you walk through the park, or do you choose an alternative route? Why?
3. **Think of a time when someone did something that you disagreed with, as it went against your moral values. How did you react? How did that make you feel?**
Scenario: You and your co-worker get along well. One day, your boss lets you know that the co-worker called in sick. This is not something that happens a

lot. Next time you chat to them, they say that they were not actually sick, but they just really needed a day off to recharge. How does that make you feel?

4. **Think of a time when you promised somebody that you would do something, but then did not want to do it anymore. Do you keep the promise, or do you go with how you feel at the time and risk disappointing somebody?**

Scenario: You promised your friend that you will come to their dinner party this evening, but you just came back from work and you're not really feeling it. You know your friend will be disappointed if you don't come. What do you do?

5. **Are you a spontaneous person, or do you like to plan everything? Think of a time when you were planning to do something, but then the plans changed suddenly. How did this make you feel?**

Scenario: You and your partner have been planning to go for a date to one of your favourite restaurants this weekend. However, on the day, your partner surprises you: they inform you that they are taking you to a different restaurant that they think you'll enjoy more. How does that make you feel?

6. **Think of a time when you were faced with a problem. Do you tend to want to try to fix the problem immediately, or do you prefer to 'sleep on it'?**

Scenario: A friend has shared a worry with you. Do you try to immediately try to come up with a solution, or do you tend to just listen and sit with the problem and the emotions to support your friend?

Emotional expression and emotional awareness

1. **Think of a time when you received some bad news in the presence others and became upset or distressed. How did you handle this situation? Did you express your emotions in any way, or did you keep them under control?**

Scenario: You are at work, and you just received a text message with some upsetting news. You become very upset and feel like crying. What do you do?

2. **Now, think about a time when you received some very exciting news in the presence of others. How did you react? Did you express your excitement to other people?**

Scenario: You are at a family dinner, and you just found out that you passed an important exam with flying colours. What do you do?

3. **Think of a time when you were asked to tell someone how you feel. How easy was it for you to identify and express your emotions?**

Scenario: You've had a difficult day, and you meet a friend. They realise something isn't quite right and asks you what happened and if you are feeling alright. What's your reaction?

4. **Think of a time when you had an argument with somebody close to you. How do you usually behave in situations like this?**

Scenario: Your partner is unhappy with you and gets emotional – for instance, starts crying or shouting. What’s your reaction?

5. **Think of a time when somebody else got upset around you. How did it make you feel? How did you react?**

Scenario: Your close colleague comes to work very upset and starts crying. What’s your reaction?

Social connectedness and intimacy with others

1. **Think of a time when you were surrounded by people you did not know. How did you feel? Did you try to get to know them? How easy was it?**

Scenario: You have been invited to your friend’s party. They have recently moved and made a lot of new friends that you don’t know and who are also at the party. Do you approach them and try to get to know them? How does this situation make you feel?

2. **Think of a time when somebody did better at something than you. How did this make you feel? Did you feel envious, or compared yourself to them?**

Scenario: You find out that your friend did better than you at an exam that you both studied hard for. How does that make you feel?

3. **Think of a time when you met somebody you seemingly get along with. How easy did you find it to warm up and open up to them?**

Scenario: You and another person have been hanging out quite a lot in the past few months. One day, they share something with you that has really upset them. You went through something similar some time ago, so you know exactly how they feel. Do you share your similar experience with them, or do you keep it to yourself?

4. **Think of a time when you were faced with a choice of either spending time on your own or spending time with others. What would you typically choose, and why?**

Scenario: It’s Saturday, and you wake up in a good mood. You have nothing important to do today. You can either relax on the sofa on your own and watch your favourite series or join your friends for a barbecue. What do you choose to do? Why?

5. **Do you consider yourself to have a good ‘social battery’? Think of a time when you have been around other people a lot in a short period of time. How does it make you feel?**

Scenario: You have been out and about with friends for the last couple of nights and had fun. Tonight, you can either relax on the sofa on your own and watch your favourite series or join your friends for a barbecue. What do you choose to do? Why?

Appendix B:
Demographic information about participants in the pilot focus group.

Variable		Total sample N = 7	OC group N = 5	UC group N = 2
Continuous				
Age (in years)	M	29.86	30.20	29.00
	SD	9.63	11.69	2.83
	Min.	23	23	27
	Max.	51	51	31
Categorical		N (%**)	N (%***)	N (%***)
Sex at birth	Female	3 (42.86)	2 (66.67)	1 (33.33)
	Male	4 (57.14)	3 (75.00)	1 (25.00)
Gender identity	Woman	3 (42.86)	2 (66.67)	1 (33.33)
	Man	4 (57.14)	3 (75.00)	1 (25.00)
Sexual orientation	Heterosexual	3 (42.86)	2 (66.67)	1 (33.33)
	Homosexual	2 (28.57)	1 (50.00)	1 (50.00)
	Bisexual	1 (14.29)	1 (100.00)	-
	Undisclosed	1 (14.29)	1 (100.00)	-
Marital status	Single	5 (71.43)	3 (60.00)	2 (40.00)
	Married/ Civil partnership/ Living with partner	1 (14.29)	1 (100.00)	-
	In a relationship, living apart	1 (14.29)	1 (100.00)	-
	Country of residence	United Kingdom	5 (71.43)	2 (28.57)
Ethnic origin	Black	1 (14.29)	1 (100.00)	-
	Mixed/multiple ethnic groups	1 (14.29)	-	1 (100.00)
	White	5 (71.43)	4 (80.00)	1 (20.00)
Education	Graduate	7 (100.00)	5 (71.43)	2 (28.57)
English	First language	6 (85.71)	4 (66.67)	2 (33.33)
	Preferred language	1 (14.29)	1 (100.00)	-

Notes. *Refers to the percentage of responses indicated in Column B of the ASC-WP questionnaire by participants in the OC group, and percentage of responses indicated in Column A of the ASC-WP questionnaire by participants in the UC group. **Refers to the percentage of the full sample. ***Refers to the percentage of all participants within the specific category.

Appendix C:
Demographic information about participants in the main study focus groups.

Variable		Total (N = 11)	Group 1 (N = 3)	Group 2 (N = 2)	Group 3 (N = 3)	Group 4 (N = 3)
Continuous						
Age (in years)	M	27.64	25.67	26.5	24.67	33.33
	SD	5.55	3.21	2.12	0.58	8.50
	Min.	22	22	25	24	27
	Max.	43	28	28	25	43
Categorical**		N (%**)	N (%***)	N (%***)	N (%***)	N (%***)
Sex at birth	Female	4 (36.36)	1 (25.00)	-	2 (50.00)	1 (25.00)
	Male	7 (63.64)	2 (28.57)	2 (28.57)	1 (14.29)	2 (28.57)
Gender identity	Woman	4 (36.36)	1 (25.00)	-	2 (50.00)	1 (25.00)
	Man	7 (63.64)	2 (28.57)	2 (28.57)	1 (14.29)	2 (28.57)
Sexual orientation	Heterosexual	10 (90.91)	3 (30.00)	1 (10.00)	3 (30.00)	3 (30.00)
	Plurisexual	1 (9.09)	-	1 (100.00)	-	-
Marital status	Single	4 (36.36)	1 (25.00)	1 (25.00)	2 (50.00)	-
	Married/Civil partnership/Living with partner	4 (36.36)	1 (25.00)	-	-	3 (75.00)
	In a relationship, living apart	3 (27.27)	1 (33.33)	1 (33.33)	1 (33.33)	-
	Country of residence	9 (81.82)	2 (22.22)	2 (22.22)	2 (22.22)	3 (33.33)
Ethnic origin	Other	2 (18.18)	-	-	1 (100.00)	-
	Black	5 (45.45)	1 (20.00)	2 (40.00)	1 (20.00)	1 (20.00)
	Mixed/multiple ethnic groups	3 (27.27)	1 (33.33)	-	1 (33.33)	1 (33.33)
	White	2 (18.18)	1 (50.00)	-	-	1 (50.00)
Education	Other	1 (9.09)	-	-	1 (100.00)	-
	Further	1 (9.09)	-	-	1 (100.00)	-
	Undergraduate	2 (18.18)	1 (50.00)	-	1 (50.00)	-
	Graduate	8 (72.72)	2 (25.00)	2 (25.00)	1 (12.50)	3 (37.50)
English	First language	9 (81.82)	2 (22.22)	2 (22.22)	2 (22.22)	3 (33.33)
	Preferred language	2 (18.18)	1 (50.00)	-	1 (50.00)	-

Notes. *Refers to the percentage of responses indicated in Column B of the ASC-WP questionnaire.

Refers to the percentage of the full sample. *Refers to the percentage of all participants within the specific category.

Appendix D:
Items rejected based on the I-CVI values < .80 for the Relevance criterion.

Domain	Item	Relevance I-CVI
LFC	I stick to the detailed plans I make.	.67
LFC	You can never be too prepared.	.50
LFC	Breaking rules is not worth the consequences.	.33
LFC	I must be the best at everything I do.	.60
LFC	I always obsess over small details.	.67
LFC	When I do someone a favour, I expect that they will also help me when I need it.	.67
LFC	Persisting at difficult tasks makes me feel worthwhile.	.33
LFC	I like to arrange things in an ordered manner (for example, the items on my desk).	.33
LFC	I feel compelled to repeatedly check everything (for example, that the doors are locked).	.67
LFC	I often feel a strong urge to fix things immediately (e.g., arising problems or tensions in relationships).	.50
LFC	Asking for help makes me feel like I am not good enough.	.67
LFC	I obsessively check the correctness of my work.	.50
LFC	Striving for perfection in everything I do makes me feel worthwhile.	.67
LFC	*I find it extremely difficult to stick to the plans I have made.	.50
LFC	I frequently worry that others see me as incompetent.	.50
LFC	I frequently wonder if I am working hard enough.	.33
LFC	Striving to be the best is more important to me than it is to most people.	.50
LFC	I prefer to know what to expect from a situation in advance.	.67
LFC	When I look at someone else's work, I frequently think that I could do better than they did.	.67
LFC	Even when a situation requires me to break a rule, I feel guilty about it.	.50
LFC	I feel more guilty about breaking rules than most people.	.50
LFC	I find it difficult to cope with unexpected changes.	.33
LFC	Exact execution of a plan assures good quality of work.	.17
LFC	Sometimes I feel like I am out of control even when other people think I am in complete control.	.50
LFC	I not only strive to be perfect, but also try to make it look easy.	.33
LFC	I never do things at the last minute.	.33
LFC	Breaking bad habits requires little effort.	.20
LFC	Sticking to a sequence (e.g., morning routine) requires little effort from me.	.50
LFC	*Things that seem to come easily to many people require a lot of effort from me.	.17
LFC	*I often struggle to stick to a simple routine.	.50
LFC	It takes very little effort for me to resist an impulse.	.67
LFC	*I easily give in to temptations.	.50
LFC	I always stick to the plans I have made.	.33

LFC	When someone else tells me they have a problem, I immediately try to find a solution to that problem.	.67
LFC	Doing things to achieve an important goal do not require much effort from me.	.20
LRO	When people criticise me, it is because they want to be like me.	.33
LRO	I am always on the lookout for potential threats.	.67
LRO	When somebody gets under my skin, I often contemplate evening the score.	.33
LRO	There is more to learn from what went wrong than from what went right.	.60
LRO	I have been told by other people that I hold myself up to rigorous standards.	.67
LRO	I am critical of other people.	.67
LRO	I am my own best critic.	.50
LRO	*If I have had a difficult day, I reward myself with something nice.	.50
LRO	I rarely feel like I have earned a reward.	.67
LRO	I always weigh up costs and benefits before I do something.	.50
LRO	I am only willing to put energy into things that are important to me.	.17
LRO	I often downplay my achievements.	.67
LRO	I only welcome critical feedback when I have asked for it.	.50
LRO	I need proof that something will happen before I get excited about it.	.33
LRO	I only ever allow myself to be spontaneous when the situation is right (e.g., on holiday).	.67
LRO	I see most things as a potential threat.	.67
LRO	I will try to convince people that I'm right even if I know I'm not.	.50
LRO	Other people doing better than me makes me feel bad about myself.	.67
LRO	*I love trying new dishes in restaurants.	.67
LRO	*I love being the centre of attention, whether it is positive or negative.	.67
LRO	*I seek out excitement.	.50
LRO	Being called out on a mistake in front of other people is my worst fear.	.67
LSC	I feel disconnected from the world.	.67
LSC	People must prove themselves to me before we can be friends.	.67
LSC	I enjoy being difficult to get to know.	.67
LSC	I often think others have unfair advantages in life.	.67
LSC	People tell me I'm difficult to read.	.50
LSC	*My face is very expressive.	.50
LSC	*I use gestures a lot when communicating with others.	.50
LSC	I have been told I do not smile often.	.50
LSC	Strangers often approach me or smile at me in the street.	.20
LSC	Comparing myself to people who are worse off often makes me feel better about myself.	.67
LSC	I compare myself to others without even realising it.	.50
LSC	*I am comfortable sitting with my emotions and trying to understand them.	.50
LSC	I only enjoy conversations with people that I connect with on an intellectual level.	.67
LSC	I strongly dislike small talk.	.67

LSC	I sometimes give others advice even if they don't ask for it.	.67
LSC	I can rarely relax in social situations.	.67
PIE	I am proud of how well I tolerate distress.	.50
PIE	I get anxious easily.	.50
PIE	I only ever feel emotions if I allow myself to.	.60
PIE	I can "turn on" and "turn off" my emotions how I please.	.50
PIE	I never let my struggles show.	.67
PIE	Thinking about feeling emotions makes me feel uneasy.	.67
PIE	I hardly ever experience extreme emotions.	.50
PIE	*I am very open about my emotions.	.67
PIE	*My emotions are a very important part of who I am.	.67
PIE	I rarely lash out.	.50
PIE	If I ever lash out, it tends to be quite explosive.	.33
PIE	If I ever lash out, it's only around people that I know very well.	.67
PIE	Admitting that I feel sad is a sign of weakness.	.67
PIE	When I am angry, I become quiet and withdrawn.	.60
PIE	I can be passive-aggressive at times.	.50
PIE	If somebody needs my help, they should communicate it clearly.	.17
PIE	People cannot expect me to know how they feel if they don't speak up about it.	.50
PIE	I can easily shake off my emotions.	.50
PIE	It is difficult for me to control anxious thoughts.	.67
PIE	I do not get angry.	.50
PIE	*To support somebody is to help them understand their emotions on a deeper level.	.33
PIE	I would generally describe my day-to-day mood as stable.	.50
PIE	I take a lot of pride in making cynical, clever jokes.	.50
PIE	*People have told me that my face is very expressive.	.50
PIE	*When something good happens, I can't wait to share the good news with everyone.	.67

Appendix E:
Demographic characteristics (categorical) of the participant samples recruited for the cognitive interviews study in the first round.

Variable	Category	<i>N</i> (%)
Residence	Botswana	1 (9.10)
	India	1 (9.10)
	UK	9 (81.82)
English language	First	10 (90.90)
	Preferred	1 (9.10)
Education	Doctoral	1 (9.10)
	Graduate	7 (63.64)
	Undergraduate	2 (18.18)
	Further	1 (9.10)
Ethnicity	Asian	1 (9.10)
	Black	8 (72.73)
	White	2 (18.18)
Marital status	Married/Civil partn./Living with partner	2 (18.18)
	Partnered, living apart	4 (36.36)
	Single	5 (45.45)
Gender	Man	8 (72.73)
	Woman	3 (27.27)
Sex	Female	4 (36.36)
	Male	7 (63.64)
Sexual orientation	Heterosexual	7 (63.64)
	Homosexual	2 (18.18)
	Plurisexual	1 (9.10)
	Other	1 (9.10)

Appendix F:
Demographic characteristics (categorical) of the participant samples recruited for the cognitive interviews study in the second round.

Variable	Category	N (%)
Residence	UK	5 (100.00)
English language	First	3 (60.00)
	Preferred	2 (40.00)
Education	Graduate	4 (80.00)
	Undergraduate	1 (20.00)
Ethnicity	Mixed/Multiple	1 (20.00)
	White	4 (80.00)
Marital status	Married/Civil partn./Living with partner	4 (80.00)
	Single	1 (20.00)
Gender	Man	3 (60.00)
	Woman	2 (40.00)
Sex	Female	2 (40.00)
	Male	3 (60.00)
Sexual orientation	Heterosexual	4 (80.00)
	Homosexual	1 (20.00)

Appendix G:
Demographic characteristics (categorical) of the total sample of participants
recruited for the EFA and CFA studies.

Variable	Category	<i>N (%)</i>
Residence	Australia	124 (10.33)
	Canada	229 (19.08)
	Ireland	56 (4.67)
	New Zealand	48 (4.00)
	UK	527 (43.92)
	USA	216 (18.00)
English language	First	1079 (89.92)
	Preferred	110 (9.17)
	Neither	10 (.83)
	Not reported	1 (.08)
Education	Doctoral	40 (3.33)
	Graduate	256 (21.33)
	Undergraduate	532 (44.33)
	Further	262 (21.83)
	Secondary	88 (7.33)
	Primary	8 (.67)
	Other	11 (.92)
	Not reported	3 (.25)
Ethnicity	Arab/ MENA	11 (.92)
	Asian	154 (12.83)
	Black	46 (3.83)
	Hispanic/ Latino	27 (2.25)
	Mixed/ Multiple	47 (3.92)
	Native American/ Alaskan	3 (.25)
	Oceanian/ Hawaiian/ Pacific Isl.	2 (.17)
	White	894 (74.50)
	Other	8 (.67)
	Not reported	8 (.67)
Marital status	Married/Civil partn./ Living with partner	646 (53.83)
	Partnered, living apart	110 (9.17)
	Separated/ divorced	65 (5.42)
	Single	360 (30.00)
	Widowed	10 (.83)
	Other	1 (.08)
	Not reported	8 (.67)
Gender	Man	464 (38.67)
	Non-binary	13 (1.08)
	Woman	713 (59.42)
	Other	6 (.50)
	Not reported	4 (.33)

Sex	Female	723 (60.25)
	Male	469 (39.08)
	Other	3 (.25)
	Not reported	5 (.42)
Sexual orientation	Asexual	12 (1.00)
	Heterosexual	993 (82.75)
	Homosexual	53 (4.42)
	Plurisexual	116 (9.67)
	Other	3 (.25)
	Not reported	23 (19.17)

Appendix H:
Differences in selected demographic characteristics (categorical) between EFA and CFA participant samples following sample split.

Variable	Category	N (EFA)	N (CFA)	χ^2	p
Education	Doctoral	17	23	28.03	.907
	Graduate	127	129		
	Undergraduate	266	266		
	Further	124	138		
	Secondary	53	35		
	Primary	5	3		
	Other	7	4		
	Not reported	1	2		
Ethnicity	Arab/ MENA	3	8	29.87	.961
	Asian	89	65		
	Black	20	26		
	Hispanic/ Latino	15	12		
	Mixed/ Multiple	20	27		
	Native American/ Alaskan	2	1		
	Oceanian/ Hawaiian/ Pacific Isl.	1	1		
	White	441	453		
	Other	6	2		
	Not reported	3	5		
Marital status	Married/Civil partn./ Living with partner	325	321	32.08	.369
	Partnered, living apart	55	55		
	Separated/ divorced	33	32		
	Single	176	184		
	Widowed	6	4		
	Other	1	-		
	Not reported	4	4		
Gender	Woman	361	352	9.38	.377
	Man	228	236		
	Non-binary	7	6		
	Other	2	4		
	Not reported	2	2		
Sexual orientation	Asexual	7	5	9.07	.928
	Heterosexual	497	496		
	Homosexual	30	23		
	Plurisexual	55	61		
	Other	2	1		
	Not reported	9	14		

Note. *Degrees of freedom were not determined due to p-values being obtained through Monte Carlo simulation.

Appendix I:
Skewness and kurtosis values for scales (and scale subscales) included in the battery of measures.

Scale	Subscales	Skewness	Kurtosis
OAQ	Global	-0.42	3.07
	LFC	-0.48	3.47
	LRO	-0.40	2.80
	LSC	-0.08	2.50
	PIE	-0.43	2.63
EUC-13	Global	-0.29	2.66
	Uninhibited behavior	0.03	2.48
	Planful conscientious behavior	-0.73	2.83
	Socially restricted behavior	-0.58	2.52
ER	-	0.05	2.87
SUPPS-P	Negative Urgency	0.11	2.35
	Lack of Perseverance	0.38	2.64
	Lack of Premeditation	0.53	3.12
	Sensation Seeking	0.34	2.99
	Positive Urgency	0.48	3.02
BIS (BAI)	-	-0.27	2.59
BAS (BAI)	Reward Responsiveness	-0.09	3.18
	Drive	0.19	2.74
	Fun Seeking	0.11	3.32
LSCM	Global	-0.02	2.72
	Impulsivity	0.25	3.05
	Simple Tasks	0.10	2.99
	Risk Seeking	0.42	2.55
	Physical Activities	0.36	2.68
	Self-Centered	0.19	2.53
	Temper	0.15	2.51
BSCS	-	0.20	2.88

Appendix J:
Demographic characteristics (categorical) of the sample used to
validate the OAQ.

Variable	Category	N (%)
Residence	Australia	65 (21.67)
	Canada	70 (23.33)
	Ireland	6 (2.00)
	New Zealand	29 (9.67)
	UK	111 (37.00)
	USA	19 (6.33)
English language	First	248 (82.67)
	Preferred	46 (15.33)
	Neither	5 (1.67)
	Not reported	1 (.33)
Education	Doctoral	14 (4.67)
	Graduate	62 (20.67)
	Undergraduate	123 (41.00)
	Further	66 (22.00)
	Secondary	28 (9.33)
	Primary	3 (1.00)
	Other	3 (1.00)
	Not reported	1 (.33)
Ethnicity	Arab/ MENA	1 (.33)
	Asian	51 (17.00)
	Black	6 (2.00)
	Hispanic/ Latino	6 (2.00)
	Mixed/ Multiple	9 (3.00)
	Native American/ Alaskan	-
	Oceanian/ Hawaiian/ Pacific Isl.	2 (.67)
	White	220 (73.33)
	Other	2 (.67)
	Not reported	3 (1.00)
Marital status	Married/Civil partner/ Living with partner	143 (47.67)
	Partnered, living apart	32 (10.67)
	Separated/ divorced	15 (5.00)
	Single	105 (35.00)
	Widowed	1 (.33)
	Other	1 (.33)
	Not reported	3 (1.00)
Gender	Man	145 (48.33)
	Non-binary	2 (.67)
	Woman	148 (49.33)
	Other	2 (.67)
	Not reported	3 (1.00)

Sex	Female	149 (49.67)
	Male	148 (49.33)
	Other	-
	Not reported	3 (1.00)
Sexual orientation	Asexual	7 (2.33)
	Heterosexual	237 (79.00)
	Homosexual	12 (4.00)
	Plurisexual	36 (12.00)
	Other	1 (.33)
	Not reported	7 (2.33)

Appendix K:
Demographic characteristics (categorical) of the sample used to examine the test-retest validity of the OAQ.

Variable	Category	N (%)
Residence	Australia	44 (23.91)
	Canada	38 (20.65)
	Ireland	5 (2.72)
	New Zealand	19 (10.33)
	UK	71 (38.58)
	USA	7 (3.80)
English language	First	155 (84.24)
	Preferred	26 (14.13)
	Neither	2 (1.09)
	Not reported	1 (.54)
Education	Doctoral	10 (5.43)
	Graduate	43 (23.37)
	Undergraduate	71 (38.59)
	Further	33 (17.93)
	Secondary	20 (10.87)
	Primary	2 (1.09)
	Other	4 (2.17)
	Not reported	1 (.54)
Ethnicity	Arab/ MENA	1 (.54)
	Asian	23 (12.50)
	Black	4 (2.17)
	Hispanic/ Latino	3 (1.63)
	Mixed/ Multiple	4 (2.17)
	Native American/ Alaskan	-
	Oceanian/ Hawaiian/ Pacific Isl.	3 (1.63)
	White	139 (75.54)
	Other	4 (2.17)
	Not reported	3 (1.63)
Marital status	Married/Civil partner/ Living with partner	94 (51.09)
	Partnered, living apart	17 (9.24)
	Separated/ divorced	9 (4.89)
	Single	61 (33.15)
	Widowed	1 (.54)
	Other	-
	Not reported	2 (1.09)
Gender	Man	90 (48.91)
	Non-binary	1 (.54)
	Woman	89 (48.37)
	Other	-
	Not reported	4 (2.17)

Sex	Female	92 (50.00)
	Male	90 (48.91)
	Other	-
	Not reported	2 (1.09)
Sexual orientation	Asexual	-
	Heterosexual	148 (80.43)
	Homosexual	5 (2.72)
	Plurisexual	20 (10.87)
	Other	5 (2.72)
	Not reported	6 (3.26)

Appendix L:
Correlation matrix of measures included in the battery of measures.

	SUPPS NU	SUPPS PE	SUPPS PR	SUPPS SS	SUPPS PU	BIS	BAS RR	BAS D	BAS FS	LSCS	LSCS I	LSCS ST	LSCS RS	LSCS PA	LSCS SC	LSCS T	BSCS	EUC-13	EUC-13 UB	EUC-13 PCB	EUC-13 SRB
SUPPS PE	.18																				
SUPPS PR	.33	.46																			
SUPPS SS	.38	-.05	.15																		
SUPPS PU	.45	.02	.28	.65																	
BIS	.34	.08	.02	-.08	-.17																
BAS RR	.16	-.17	-.05	.28	.21	.07															
BAS D	.17	-.29	.00	.33	.37	-.25	.48														
BAS FS	.24	.03	.26	.54	.57	-.20	.41	.47													
LSCS	.57	.11	.38	.44	.52	-.04	.16	.28	.49												
LSCS I	.46	.28	.44	.26	.36	-.03	.08	.05	.37	.67											
LSCS ST	.30	.30	.17	-.13	-.08	.38	-.10	-.25	-.14	.38	.29										
LSCS RS	.30	.03	.29	.59	.62	-.31	.19	.39	.62	.66	.43	-.13									
LSCS PA	.08	-.17	.08	.31	.35	-.19	.23	.29	.46	.48	.14	-.11	.43								
LSCS SC	.28	-.07	.09	.25	.31	-.19	.02	.25	.19	.57	.26	.09	.33	.11							
LSCS T	.54	-.01	.22	.15	.16	.25	.08	.14	.11	.61	.25	.31	.15	.05	.33						
BSCS	-.54	-.39	-.47	-.26	-.31	-.27	-.09	.01	-.28	-.49	-.60	-.39	-.28	.13	-.19	-.32					
EUC-13	.31	.20	.50	.32	.46	-.29	.06	.25	.46	.46	.38	-.01	.53	.22	.24	.19	-.32				
EUC-13 UB	.38	.09	.27	.44	.55	-.20	.17	.34	.49	.55	.43	-.03	.56	.22	.35	.30	-.39	.67			
EUC-13 PCB	.17	.32	.50	.05	.18	-.05	-.14	-.07	.17	.21	.22	.14	.20	.01	.06	.10	-.26	.68	.13		
EUC-13 SRB	.02	-.07	.14	.15	.18	-.38	.11	.30	.28	.13	.04	-.20	.28	.27	.05	-.03	.11	.60	.23	.14	
ER-89	-.17	-.22	-.14	.26	.26	-.46	.27	.37	.44	-.02	-.10	-.43	.35	.33	-.07	-.21	.27	.22	.21	-.06	.36

Appendix M:
The final version of the OAQ with participant and scoring instructions.

Overcontrolled Assessment Questionnaire (OAQ)

Participant instructions

Please review the instructions carefully before filling out the questionnaire. Below you will see a list of questions about **how you typically behave and respond to situations in everyday life**. There are seven possible answers for each question: (1) Strongly disagree, (2) Disagree, (3) Slightly disagree, (4) Slightly agree, (5) Agree, (6) Strongly agree, and (7) No opinion. **Please read each question and select the answer that best describes you as a person.** While you may not find a response that exactly reflects who you are, try to pick the one that is the closest to how you would describe yourself. There are no right or wrong answers – try to go with the answer that first comes to mind.

Scale items

1. I like to be prepared for any possibility.
2. I make detailed plans for how to do things.
3. I feel the need to fix problems immediately.
4. Structure and order in everyday life are important to me.
5. I like my life well-structured and predictable.
6. I cannot stop thinking about a problem until I find a solution.
7. I avoid situations in which I could be criticised.
8. The world is a scary place.
9. When I make a mistake in front of others, I feel more embarrassed than most people would.
10. I often see a threat where others may not.
11. Embarrassing myself in front of others is one of my worst fears.
12. When someone gives me a compliment, I ask myself why they are complimenting me.
13. I rarely feel deeply connected to other people.
14. I often struggle to understand another person's perspective.
15. My relationships with other people are rather shallow.
16. When other people become upset, I don't know how to support them.
17. I dislike being around people who are upset.
18. When I see another person crying, I find it difficult to understand why they would do this in public.
19. I am distant from other people.
20. I tend to dismiss my emotions.
21. I am an open book to other people. **(R)**
22. I push through difficult situations without sharing my struggles with others.
23. I don't reveal my vulnerability.
24. Other people express emotions more openly than I do.
25. No matter what I feel on the inside, I make sure I seem fine on the outside.
26. I tend to bottle up my feelings.

Information for administrators

This four-dimensional scale has been developed by based on the Neurobiosocial Theory for Disorders of Overcontrol (Lynch, 2018) to measure the construct of maladaptive overcontrol.

Items 1-6 pertain to the *low flexible control* (LFC) domain.

Items 7-12 pertain to the *low responsivity and openness* (LRO) domain.

Items 13-19 pertain to the *low social connectedness and intimacy with others* (LSC) domain.

Items 20-26 pertain to the *pervasive emotional inhibition and low emotional awareness* (PIE) domain.

Items should be presented to participants in a random order.

Responses options are scored as follows:

Strongly disagree – 1 point,

Disagree – 2 points,

Slightly disagree – 3 points,

Slightly agree – 4 points,

Agree – 5 points,

Strongly agree – 6 points,

No opinion – n/a; the response should be treated as missing data.

The extent of difficulties pertaining to each deficit is indicated by the corresponding total score.

The extent of maladaptive overcontrol is indicated by total global score.

Question 21 must be reverse-scored prior to calculating the totals.

Higher total scores indicate higher levels of maladaptive overcontrol.
