The Development and Implementation of Youth Mental Health Outcome Measures within Measurement Feedback Systems

Benjamin Kwan

University of Canberra

Faculty of Health

Primary Supervisor: Professor Debra J. Rickwood

Secondary Supervisor: Dr Patricia M. Brown

A thesis by published works submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Clinical Psychology

12th April 2021
Abstract

This thesis by published works provides an original contribution to the knowledge about mental health outcome measures used with young people aged 12 to 25 years. It discusses how such measures can be developed and implemented as a clinical tool in day-to-day practice. The main goal was to develop a routine outcome measure that was suitable for youth mental health settings that could be implemented into measurement feedback systems. The research aimed to achieve this goal by firstly identifying the gap in existing mental health outcome measures used with young people and, secondly, by examining how these measures were being used clinically. A further aim of this research was to examine the psychometric properties of a new routine outcome measure, MyLifeTracker, in youth mental health settings. It aimed to determine developmentally appropriate clinically significant change indexes, expected change trajectories, and early warning signals for MyLifeTracker, to provide clinicians with clinically useful and evidence-based benchmarks. Lastly, the research aimed to explore factors affecting the use of MyLifeTracker in measurement feedback systems across youth mental health settings.

The research used a quantitative methodology comprising four research papers, each addressing a specific research aim. The first research paper was a systematic review of general mental health outcome measures for young people aged 12 to 25 years. It identified how these measures track change and if they had been used in feedback monitoring. The second and third papers examined the use of MyLifeTracker by clients receiving support from headspace youth mental health services across Australia. The second paper explored the reliability, validity, and sensitivity to change of MyLifeTracker for young people, across gender and age groups. The third paper determined clinically significant change indexes by gender and age groups by comparing participants from headspace services to an Australian representative community sample of young people. Expected change trajectories were also
determined for the clinical group using growth curve modelling. The fourth paper reports findings from a survey of 210 clinicians from *headspace* centres about their use of MyLifeTracker, specifically exploring three processes of measurement feedback systems: looking at MyLifeTracker before the session, using MyLifeTracker in treatment planning, and providing feedback of MyLifeTracker scores to clients.

The systematic review identified 29 different outcome measures used with young people, however, no measures were explicitly designed for this age group. Only two measures were found to be used by clinicians in measurement feedback systems in this age range. Findings from the review led to the recommendation that measures be explicitly designed for this age group that are suitable for routine outcome monitoring. The second paper demonstrates that MyLifeTracker provides a psychometrically sound mental health outcome measure for routine use with young people. The measure has been incorporated into an electronic system for *headspace* services that routinely tracks session-by-session change and produces time-series charts for ease of use and interpretation. The third paper provides clinical benchmarks for MyLifeTracker, further supporting the use of the measure in measurement feedback systems. Lastly, the fourth paper reports the different levels of use of MyLifeTracker in a measurement feedback system and highlights the factors that increase clinicians’ use for each process.

The thesis supports the use of mental health outcome measures to be used not only for assessing service effectiveness and quality assurance, but also as a clinical tool to support decision making and treatment planning by clinicians and clients. It provides support for brief and easy to use measures that are meaningful for clinicians and clients. The clinical benefits of measurement feedback systems are becoming more widely known and have become part of the agenda for the future progression of psychotherapy. The thesis targets a developmental age group that has high rates of clinical deterioration, treatment drop-out and missed
appointments, and who may be quite responsive to feedback monitoring. The thesis concludes by offering a range of targeted strategies that can support the future implementation of outcome measurement feedback systems into practice.
### List of Included Published Works

**Paper 1**  
https://doi.org/10.1186/s12888-015-0664-x

**Paper 2**  

**Paper 3**  
https://doi.org/10.1111/eip.13016

**Paper 4**  
Conference Presentations During PhD Candidature

**Conference 1**

**Conference 2**
Acknowledgements

First and foremost, I want to thank my primary supervisor, Professor Debra Rickwood. The support, expertise, and patience you have provided throughout this time has been invaluable. The passion and commitment to your work is something I aspire to.

I would also like to thank my secondary supervisor, Dr Tricia Brown, for the support during my honours year and the help during the final stages of this research project. I am grateful for the calm presence and knowledge you bring to your work.

Dr Kerry Thomas of Proof, Edit & Style provided professional editing services in accordance with the Institute of Professional Editors’ Guidelines for editing research theses.

I would like to thank my family and friends for always being there and supporting me throughout my life. Thank you to my fellow students from the University of Canberra psychology program, the friendship and shared learning over the years is something I will remember. A big thank you to my colleagues from headspace, the encouragement and guidance has been much appreciated.

Finally, I would like to thank my amazing partner, Sarah, and our young pup, Hugo. I am eternally grateful for the unconditional love and support you have provided during this time, and I’m excited about the next adventures we have together!
Table of Contents

Certificate of Authorship of Thesis........................................................................................................ iii

Abstract.................................................................................................................................................. v

List of Included Published Works........................................................................................................ ix

Conference Presentations During PhD Candidature........................................................................ xi

Acknowledgements................................................................................................................................ xiii

Table of Contents.................................................................................................................................. xv

List of Tables ........................................................................................................................................ xxi

List of Figures ........................................................................................................................................ xxiii

List of Appendices ............................................................................................................................. xxv

List of Abbreviations ........................................................................................................................... xxvii

Chapter I: Introduction........................................................................................................................... 1

1.1 Significance..................................................................................................................................... 1

1.2 Rationale ......................................................................................................................................... 1

1.3 Aims ............................................................................................................................................... 4

1.4 Thesis Structure ............................................................................................................................. 5

Chapter II: Literature Review ............................................................................................................... 7

2.1 Youth Mental Health and Wellbeing ............................................................................................. 7

2.1.1 Defining “young people” ........................................................................................................ 7

2.1.2 Prevalence of mental health problems .................................................................................. 9

2.1.3 Impact of mental health problems .......................................................................................10
2.2 Youth Mental Health Care ............................................................................................. 11

  2.2.1 Help-seeking and engagement ................................................................................ 11

  2.2.2 Youth-focused services .......................................................................................... 14

2.3 Mental Health Outcome Measurement .......................................................................... 16

  2.3.1 Broad mental health outcome measures ................................................................. 16

  2.3.2 Domains of mental health outcome measures ....................................................... 17

  2.3.3 Session-by-session outcome measures ................................................................... 20

2.4 Mental Health Outcome Measures for Young People ................................................... 22

  2.4.1 Development of outcome measures for young people ............................................ 22

  2.4.2 Session-by-session outcome measures for young people ....................................... 25

  2.4.3 Electronic administration of outcome measures .................................................... 26

  2.4.4 Developmentally specific domains of outcome measures for young people ......... 28

2.5 Tracking Change Using Outcome Measures ................................................................ 30

  2.5.1 Paired t-test and effect sizes .................................................................................... 30

  2.5.2 Reliable change indexes and clinically significant change indexes ......................... 31

  2.5.3 Growth curves (expected change trajectories) ........................................................ 34

2.6 Measurement Feedback Systems ................................................................................... 38

  2.6.1 Clinician blindspots ................................................................................................ 38

  2.6.2 Early warning systems in measurement feedback systems ..................................... 42

  2.6.3 Shared decision making and treatment planning ..................................................... 44

2.7 Effectiveness of Measurement Feedback Systems ........................................................ 46
2.7.1 Measurement feedback systems with at risk populations .................................................. 48
2.7.2 Clinician fidelity in measurement feedback systems ..................................................... 51
2.8 Implementation of Measurement Feedback Systems .......................................................... 52
2.8.1 Clinician characteristics and attitudes .......................................................................... 52
2.8.2 Organisational supports ............................................................................................... 55
2.8.3 Processes and fidelity ..................................................................................................... 56
2.8.4 Implementation with young people ............................................................................. 57
2.9 Chapter Summary and Rationale for Current Research .................................................... 58
2.10 The Implementation of MyLifeTracker in headspace .................................................... 59
2.11 Research Aims .................................................................................................................. 61

Chapter III: A Systematic Review of Mental Health Outcome Measures for Young People Aged 12 to 25 Years .................................................................................................................. 67

Declaration of co-authored publication Chapter III ............................................................... 68

Chapter IV: Development and Validation of MyLifeTracker: A Routine Outcome Measure for Youth Mental Health .......................................................................................................................... 89

Declaration of co-authored publication Chapter IV ............................................................... 90

Chapter V: A Routine Outcome Measure for Youth Mental Health: Clinically Interpreting MyLifeTracker .................................................................................................................................................. 103

Declaration of co-authored publication Chapter V ............................................................... 104

Chapter VI: Factors Affecting the Implementation of an Outcome Measurement Feedback System in Youth Mental Health Settings .................................................................................................................. 117

Declaration of co-authored publication Chapter VI ............................................................... 118
7.5.1 Measurement feedback system implementation across different settings .................172
7.5.2 Outcome measurement in organisational decision making .....................................174
7.6 Strengths and Limitations ............................................................................................177
7.7 Future Research Directions ..........................................................................................181
  7.7.1 Measurement feedback systems with young people .............................................181
  7.7.2 Clinical practice implications ..............................................................................185
  7.7.3 Clinician development implications ......................................................................186
  7.7.4 Organisational implications ................................................................................187
7.8 Conclusion ..................................................................................................................187
References .......................................................................................................................191
Appendices .....................................................................................................................237
List of Tables

Chapter III Tables
Table 1 Search terms 71

Chapter IV Tables
Table 1 Descriptive statistics for MyLifeTracker and standardized outcome measures at first headspace visit, by age group and gender 96
Table 2 Pearson correlation coefficients between MyLifeTracker and standardized outcome measures, by age group and gender 96
Table 3 Proportion of young people showing reliable change for MyLifeTracker between first and last session ratings, by age group and gender 97
Table 4 Proportion of young people showing reliable change for K10 between first and last session ratings, by age group and gender 97

Chapter V Tables
Table 1 Descriptive statistics for MyLifeTracker for the clinical and nonclinical groups, and clinically significant change indexes, by age group and gender 109
Table 2 Growth curve models for MyLifeTracker scores during treatment by baseline score 110
Table 3 Summary of clinical benchmarks for MyLifeTracker 112

Chapter VI Tables
Table 1 Participant demographic characteristics 122
Table 2 Subscale scores for MFA, AM-MF, and OFM, and their effect sizes when compared to a neutral rating of 3

Table 3 Intercorrelations among predictor and dependent variables

Table 4 Standardised multiple regression coefficients (β) for predicting use of MLT in MFS: before session ($n = 161$), treatment planning ($n = 159$), and client feedback ($n = 157$)
List of Figures

Chapter III Figures

Figure 1 PRISMA flow diagram 72
Figure 2 Age range and mean of cognition and emotion measures 73
Figure 3 Age range and mean of functioning measures 74
Figure 4 Age range and mean of quality of life measures 75
Figure 5 Age range and mean of multidimensional measures 76

Chapter IV Figures

Figure 1 Average rate of change in MLT across weeks in treatment 98

Chapter V Figures

Figure 1 Expected change trajectories by baseline MyLifeTracker scores 109
Figure 2 MyLifeTracker graph with benchmarks and early warning signals 110
for a 15 to 17 year old female with a baseline score between 21 and 30
Figure 3 MyLifeTracker graph with benchmarks and early warning signals 111
for a 12 to 14 year old female with a baseline score between 31 and 40
List of Appendices

Appendix A  Paper One PRISMA 2009 Checklist  237
Appendix B  Paper One Mental Health Outcome Measures Used With Young People  239
Appendix C  Papers Two and Three Letter from headspace National for Use of Data  247
Appendix D  Papers Two and Three UC Ethics Approval  249
Appendix E  Paper Four UC Ethics Approval  251
Appendix F  Paper Four Recruitment Email  253
Appendix G  Paper Four Participant Information and Consent Form  255
Appendix H  Paper Four Questionnaire  257
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMHS</td>
<td>Adult Mental Health Services</td>
</tr>
<tr>
<td>CAMHS</td>
<td>Child and Adolescent Mental Health Services</td>
</tr>
<tr>
<td>CBT</td>
<td>Cognitive Behavioural Therapy</td>
</tr>
<tr>
<td>CORE-10</td>
<td>Clinical Outcomes in Routine Evaluation-10</td>
</tr>
<tr>
<td>COVID-19</td>
<td>Coronavirus Disease 2019</td>
</tr>
<tr>
<td>CSI</td>
<td>Clinically Significant Indexes</td>
</tr>
<tr>
<td>GAD-7</td>
<td>Generalised Anxiety Disorder-7</td>
</tr>
<tr>
<td>GHQ-12</td>
<td>General Health Questionnaire-12</td>
</tr>
<tr>
<td>GP</td>
<td>General Practitioner</td>
</tr>
<tr>
<td>HoNOS</td>
<td>Health of the Nation Outcome Scale</td>
</tr>
<tr>
<td>HoNOSCA</td>
<td>Health of the Nation Outcome Scale for Children and Adolescents</td>
</tr>
<tr>
<td>K10</td>
<td>Kessler Psychological Distress Scale 10</td>
</tr>
<tr>
<td>LGBTIQ+</td>
<td>Lesbian, Gay, Bisexual, Transgender/Gender Diverse, Intersex and Queer</td>
</tr>
<tr>
<td>MFS</td>
<td>Measurement Feedback Systems</td>
</tr>
<tr>
<td>MLT</td>
<td>MyLifeTracker</td>
</tr>
<tr>
<td>ORS</td>
<td>Outcome Rating Scale</td>
</tr>
<tr>
<td>OQ-45</td>
<td>Outcome Questionnaire-45</td>
</tr>
<tr>
<td>PCOMS</td>
<td>Partners for Change Outcome Management System</td>
</tr>
</tbody>
</table>
PHQ-2 Patient Health Questionnaire-2
PHQ-9 Patient Health Questionnaire-9
RCI Reliable Change Indexes
RCT Randomised Controlled Trial
SDM Shared Decision Making
SDQ Strengths and Difficulties Questionnaire
SxS-SDQ Session-by-Session Strengths and Difficulties Questionnaire
TAU Treatment as Usual
Y-OQ Youth Outcome Questionnaire
YP-CORE Young Person-Clinical Outcomes in Routine Evaluation
Chapter I: Introduction

1.1 Significance

This thesis by published works makes an original contribution to the knowledge and understanding of mental health outcome measures used with young people aged 12 to 25 years and the implementation of such measures in measurement feedback systems (MFS). An overview of general mental health measures used with young people is provided and furthers the knowledge about their suitability and limitations for a range of settings and populations. The thesis provides a framework for the development and validation of a new session-by-session mental health outcome measure for young people that can be suitable for a range of presentations while being clinically useful. The development of structured supports and statistically derived benchmarks that enhance and optimise a measure’s ability to be used in MFS is discussed. This includes the determination of clinically significant indexes (CSI), reliable change indexes (RCI), expected change trajectories, and early warning signals. Additionally, the thesis includes clinical examples and a summary guide of the benchmarks, which deepens the understanding around MFS use with young people. Lastly, further research into factors affecting the implementation of MFS in youth mental health settings is provided, with an in-depth investigation of the different factors that affect the specific processes of MFS use. Overall, the thesis provides further knowledge around the development and implementation of outcome measures used in MFS for youth mental health settings. This is a shift away from outcome measures being used solely as an instrument for evaluating service delivery and quality assurance.

1.2 Rationale

Young people have the highest burden of mental illness across the lifespan, comprising 45% of the burden of disease for young people aged 10 to 24 years (Gore et al.,
2011). At least one in four young people aged 12 to 24 years experience a mental health problem in any given year (T. Slade, Grove, & Burgess, 2011). Young people have higher rates of deterioration compared to adults in mental health treatment (Finch, Lambert, & Schaalje, 2001; Warren, Nelson, & Burlingame, 2009) and clinicians have lower rates of accurately predicting significant decline for young people, compared to predicting decline for adults receiving treatment (Lambert et al., 2002; Nelson, Warren, Gleave, & Burlingame, 2013). Young people are also more likely to show higher treatment drop-out and missed appointments, and this could be due to their perceptions around the usefulness of professional help and stigma related to this (O’Brien, Fahmy, & Singh, 2009).

Traditional mental health services are organised around a demarcation between Child and Adolescent Mental Health Services (CAMHS) and a significant transition to Adult Mental Health Services (AMHS) at the age of 18 years (Cleverley, Lenters, & McCann, 2020). Mental health reform in Australia has reoriented mental health services to focus on early intervention for young people aged 12 to 25 years, and this approach has been gaining momentum in other countries (McGorry, Bates, & Birchwood, 2013). Historically, outcome measures have either been targeted towards children and adolescents or adults, reflecting the traditional demarcation within the mental health care system. For example, a suite of measures used in the Australian public mental health system has two versions: the Health of the Nation Outcome Scale (HoNOS) for adults aged 18 to 64 years, and the Health of the Nation Outcome Scale for Children and Adolescents (HoNOSCA) for those aged under 18 years (Pirkis et al., 2005). Outcome measures specific to the youth transition period of adolescence and young adulthood are urgently needed due to recent changes in mental health service delivery explicitly targeting this age range (McGorry et al., 2013).

There is an expanding implementation of session-by-session measures, whereby clients complete outcome measures at each session to quantitatively inform clinicians about
their progress (Hall et al., 2015). Brief measures with very few items appeal to services in which regular outcome monitoring is required and where clinicians want to utilise measures in MFS. The use of MFS, also known as Feedback Informed Therapy and Routine Outcome Monitoring, describes a process where a clinician receives feedback on a client’s progress through routine outcome measures that can be discussed within the session to help guide treatment planning (Miller, Hubble, Chow, & Seidel, 2015). The inclusion of change indexes, such as CSI, RCI, and expected change trajectories, allow clinicians to compare clients’ progress against statistically derived benchmarks (Jacobson & Truax, 1991; Singer & Willett, 2003).

MFS can improve communication between client and clinician, increase the accuracy of diagnosis, and enable rapid adjustments to treatment planning when required (Carlier et al., 2012). Furthermore, MFS can support more substantial outcome effects and improve the efficiency of treatment (Bickman, Kelley, Breda, de Andrade, & Riemer, 2011; Janse, De Jong, Van Dijk, Hutschemaekers, & Verbraak, 2017). MFS are particularly useful for clients not on track or who are at risk of treatment failure, with feedback systems significantly reducing deterioration rates and increasing rates of clinically significant improvement (Lambert, Whippe, & Kleinstäuber, 2018; Shimokawa, Lambert, & Smart, 2010). Specifically, deterioration rates of up to 21% have been seen in youth psychotherapy settings and it is recommended that outcome measurement systems that include early warning signals are implemented in such settings (Warren et al., 2009). Notably, the use of MFS across youth mental health settings could reduce the difficulties with engagement and high drop-out rates seen with young people (Donald, Carey, & Rickwood, 2018; Langer & Jensen-Doss, 2018; Smith & Jensen-Doss, 2017).

The benefits of MFS are becoming more widely known and there is increased training in the use of MFS in clinical programs (Overington, Fitzpatrick, Hunsley, & Drapeau, 2015).
However, clinicians still have a typically low completion rate of routine outcome measures, which suggests difficulties with the clinical use of MFS (Batty et al., 2013; Hatfield & Ogles, 2004; Johnston & Gowers, 2005). Several clinician characteristics have been shown to increase MFS use, or bring about more positive attitudes towards MFS use, such as being a Cognitive Behavioural Therapy (CBT) practitioner, fewer years of clinical experience, clinicians with higher-level degrees, and those in institutional settings (Ionita & Fitzpatrick, 2014; Jensen-Doss et al., 2018). Clinician attitudes towards outcome measures and MFS have also been shown to affect the amount of MFS use. Clinicians who are more positive about a measure’s clinical utility, treatment planning properties, and practicality tend to use routine outcome measures more frequently. Service and organisational factors also play a vital role in the successful implementation of routine outcome measures and MFS. Important factors that have been identified include practical resources, championing, clinical support, and training (Bickman, Douglas, et al., 2016; Gleacher et al., 2016).

Krägeloh, Czuba, Billington, Kersten, and Siegert (2015) have proposed several processes within MFS: clinician-only feedback, feedback to clinicians and clients, encouragement of mutual measurement discussion, and the availability of formalised mechanisms to guide such discussion. Clinicians’ fidelity to using all the MFS processes has been shown to affect client treatment outcomes (De Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012). Further exploration into the factors affecting the implementation of MFS in these processes is justified to provide an in-depth understanding of barriers and facilitators to MFS use.

1.3 Aims

The overarching goal of this research was to explore mental health outcome measures used with young people aged 12 to 25 years and determine a routine outcome measure that
was suitable for youth mental health settings that could be implemented within MFS. The research aimed to achieve this goal by identifying the gap in existing mental health outcome measures used with young people and examining how these measures were being used clinically. The research aimed to examine the psychometric properties of the MyLifeTracker (MLT) measure, a routine outcome measure for youth mental health settings that was implemented in *headspace* youth mental health centres throughout Australia. A further aim was to determine developmentally appropriate evidence-based benchmarks for this measure, to provide clinicians with clinically useful information. Lastly, the research aimed to explore factors affecting the measure’s use within MFS across youth mental health settings.

### 1.4 Thesis Structure

This thesis is submitted in the format of a thesis by published works. The research used a quantitative multiphase design, which included four distinct studies. Following the introduction (Chapter I), the thesis presents an overview of the current literature (Chapter II). This literature review summarises the research related to mental health outcome measurement with young people and its uses in MFS. Initially, it provides a background on young people’s mental health and wellbeing, and the current state of mental health care for this population. The literature review introduces the range of broad mental health outcome measures, discussing their different qualities and uses, particularly those used with young people. The literature review then highlights MFS, exploring the benefits and effectiveness of their use, and the factors affecting the implementation of such systems. Chapter III is a systematic review of general mental health outcome measures for young people aged 12 to 25 years. This is followed by Chapter IV, which examines the development and validation of a youth mental health outcome measure, MLT. Chapter V provides statistically derived benchmarks for MLT for clinical use. Chapter VI is a study exploring the factors affecting the
implementation of MLT within MFS in youth mental health settings. Finally, Chapter VII presents a synthesis of the research findings and includes a discussion of the practical implications for MFS in youth mental health settings. It also provides a discussion on the research limitations, future directions for research, and overall conclusions.
Chapter II: Literature Review

This chapter introduces the literature related to youth mental health, routine outcome measurement, and the implementation of MFS. Initially, a background will be provided around the current state of youth mental health and the therapeutic supports provided in this area. The review will introduce outcome measurement and the range of domains that cover the mental health spectrum. A discussion is provided about the development of session-by-session measures, electronic measurement systems, outcome measures for young people, and the differing methods for tracking change. This will lead to the introduction of MFS and its benefits, which include reducing clinician blindspots, aiding shared decision making (SDM), and supporting treatment planning. The evidence around the effectiveness of MFS will be discussed, exploring a range of outcome measurement systems, mental health settings, and client populations. This will be supplemented with research focusing on accuracy levels for detecting therapeutic failure using MFS, using MFS with young people, and clinician fidelity using MFS. The factors affecting the implementation of MFS are summarised, focusing on clinician characteristics, clinician attitudes, organisational supports, and the implementation of MFS with young people. The chapter concludes with the rationale for, and aims of, the current research.

2.1 Youth Mental Health and Wellbeing

2.1.1 Defining “young people”. Adolescence is a developmental period that is recognised by the onset of puberty and characterised by several complex changes socially, physically, neurologically, and psychologically (Keshavan, Giedd, Lau, Lewis, & Paus, 2014). These multifaceted changes continue into early adulthood, whereby there is an increase in adult roles and responsibilities, and significant transitions in education, employment, independence, and relationships (Arnett, Žukauskienė, & Sugimura, 2014).
There is an increasing shift away from the historical categorisation of adulthood starting at the age of 18 years, and an increased focus on these periods of adolescence and young adulthood that have been collectively termed “young people” (McGorry, Goldstone, Parker, Rickwood, & Hickie, 2014). The term “young people” and “youth” will be used throughout the thesis to refer to this combined period of adolescence and young adulthood, aged from 12 to 25 years.

Many structural and functional brain changes occur during this developmentally distinct period. Brain size reaches its peak during adolescence, however, there are continuing changes across the white matter, cortical, and cerebellar grey matter volume (Keshavan et al., 2014). Specifically, the pre-frontal cortex, which is vital for decision making, planning, and reasoning, is one of the last regions of the brain to fully develop. There is an increased hormone release that is responsible for sexual maturation, height growth, maturation of organ systems, and changes in the regulation of oxytocin and vasopressin, which affect attachment and social interaction (Patton & Viner, 2007). This neurodevelopment continues into early adulthood, particularly in brain regions associated with emotional and behavioural regulation (Spear, 2000).

With the numerous neurological and cognitive changes, young people show an increase in social and identity exploration, whereby they start to question who they want to be, what they want to do, and whom they want to be with (Arnett et al., 2014). Social interactions with peers become particularly important, with most adolescents talking to peers for one-third of their waking day, compared to a much lower rate with adults (Spear, 2000). It has been suggested that this helps increase social skills, which are essential in the transition to independence (Arnett et al., 2014). Risk-taking and sensation-seeking are prominent features in adolescence, with increases in misconduct, illicit substance use, and anti-social behaviours (Patton & Viner, 2007). There is instability around romantic relationships and employment,
with high levels of casual sex, short-term romantic relationships, and frequent job changes occurring in this period (Arnett et al., 2014). There are also feelings of being in-between adolescence and adulthood, with questions around the capacity to make independent decisions and being responsible for one’s self. The complex changes in physiology, emotional functioning, cognitive capacity, and significant life transitions during this period have been shown to increase the vulnerability of young people to mental health and behavioural issues (McGorry et al., 2014).

2.1.2 Prevalence of mental health problems. Young people have the highest burden of mental illness across the lifespan, comprising 45% of the burden of illness for young people aged 10 to 24 years (Gore et al., 2011). At least one in four young people aged 12 to 24 years experience a mental health problem in any given year (T. Slade et al., 2009). There is an indication that 75% of people experiencing a psychiatric disorder in adulthood experience onset of symptoms by the age of 24 years (Kessler et al., 2005). The three most prevalent mental disorders seen among Australians aged 16 to 24 years are anxiety disorders (15%), substance use disorders (13%), and affective disorders (6%) (Australian Institute of Health and Welfare, 2011). While young women show a higher prevalence of mental disorders overall, they are also more likely to have an anxiety disorder or affective disorder than young men, whereas young men are more likely to have a substance use disorder. While mental health literacy and engagement in support services have increased, the prevalence of mental health issues among young people remains high (Goodsell et al., 2017).

The high levels of mental health issues among young people coincide with the high prevalence of deliberate self-harm. Although there is significant variability among reported rates of self-harm, there is evidence to suggest that the mean lifetime prevalence rates among adolescents internationally are 16% to 18% (Muehlenkamp, Claes, Havertape, & Plener, 2012). These numbers increase to 31% when multi-item assessments for self-harm are used.
Females are more likely to self-harm than males, and self-harm is particularly common in females aged 15 to 24 years. Self-harm is associated with a range of mental disorders, alcohol and other drug use, and, most concerningly, suicide (Paul Moran et al., 2012). Globally, the second most common form of death for young people is suicide, and for females aged 15 to 19 years it is the most common form of death (Hawton, Saunders, & O'Connor, 2012). There are an estimated 164,000 self-inflicted deaths per year across the globe for those under the age of 25 years. Generally, male adolescents are more likely to die by suicide than female adolescents, occurring about two and a half times more often.

2.1.3 Impact of mental health problems. Mental health issues among young people have a significant impact on their quality of life and functioning. The adverse effects of mental health issues in young people typically spread to future social and economic outcomes (Patel, Flisher, Hetrick, & McGorry, 2007). Mental disorders amongst adolescents are related to a substantial number of days absent from schooling (Lawrence et al., 2016). Adolescents with a mental disorder are academically behind their cohort and by Year 9 they are 1.5 to 2.8 years behind those that do not have a mental disorder (Goodsell et al., 2017). Specifically, severe depression in adolescents is associated with increased comorbid mental health issues, significant role impairment, and twice as many days out of role when compared to those with mild to moderate depression (Avenevoli, Swendsen, He, Burstein, & Merikangas, 2015).

Young adults between the ages of 18 and 25 years who experience a psychiatric disorder are significantly more likely to experience lower living standards, lower income, and reduced workforce participation by the time they are 30 years old (Gibb, Fergusson, & Horwood, 2010). The rates of engagement in employment and education among young people with mental health issues are especially low for Aboriginal or Torres Strait Islanders, males, the homeless, and those using illicit substances (Holloway et al., 2018). Adolescents who experience mental health issues are more likely to use cigarettes, alcohol, and illegal
substances (Lawrence et al., 2016). The use of a combination of illicit substances during adolescence is associated with substance use disorders in young adulthood (Moss, Chen, & Yi, 2014). Young people aged 18 to 24 years are the most likely age group to use illicit substances and exceed alcohol consumption risk guidelines (Australian Institute of Health and Welfare, 2011). Those that have self-harmed during adolescence are more likely to experience a number of psychosocial issues, such as unemployment, financial hardship, and social disadvantage, as well as cigarette and illicit substance use, by the time they are 35 years old (Borschmann et al., 2017).

In 2009, the financial cost of mental illness in young people aged 12 to 25 years in Australia was $10.6 billion, and the majority of this was related to loss of productivity due to lower employment, absenteeism, and premature death (Access Economics Pty Limited, 2009). Additionally, the value of disability and premature death was estimated to be a further $20.5 billion. Of this, the Australian Government bears just over 50% of the costs related to mental illness in young people. The total economic loss of youth suicide in Australia was estimated at $511 million a year (Kinchin & Doran, 2018). Given the societal financial impacts of mental illness in young people, an investment in youth mental health care is proposed to be economically beneficial (Hamilton et al., 2017).

2.2 Youth Mental Health Care

2.2.1 Help-seeking and engagement. Typically, young people show low levels of help-seeking for mental health issues. A systematic review exploring help-seeking in youth mental health identified several substantial facilitators and barriers (Gulliver, Griffiths, & Christensen, 2010). The most prominent barrier was perceived stigma and embarrassment, and this was the primary barrier across different mental health conditions, such as depression, social phobia, and psychosis. Perceived stigma and embarrassment were also the main
impediments for accessing both formal and informal supports, such as general practitioners (GPs), counsellors, friends, and family (Yap, Reavley, & Jorm, 2013). Young people show problems in recognising symptoms of mental illness, prefer relying on themselves rather than seeking external support, and have a pervasive belief that seeking assistance does not help (Gulliver et al., 2010). More specifically, young people find it challenging to identify disorders, such as social phobia, and are less likely to seek treatment for such disorders. At the same time, depression is commonly over-generalised in young people (Reavley & Jorm, 2011). One study showed that young people have a distrust and lack of confidence towards GPs and believe that GPs are not equipped to deal with mental health issues (Leavey, Rothi, & Paul, 2011). Instead, young people are more likely to approach their family for support over professional treatment (Reavley & Jorm, 2011; Yap et al., 2013).

There is growing research into facilitators for help-seeking in youth mental health and some identified themes include past positive experiences using mental health services, encouragement by social supports to access services, and positive relationships and trust with service staff (Gulliver et al., 2010). It is suggested that past positive experiences may help increase mental health literacy, which is crucial in the help-seeking process. Specific attention has focused on populations that have lower rates of help-seeking and higher rates of mental illness, including young people who identify as Aboriginal or Torres Strait Islanders, LGBTIQ+, culturally and linguistically diverse, homeless, or male (Brown, Rice, Rickwood, & Parker, 2016; Rice, Telford, Rickwood, & Parker, 2018). Among Aboriginal or Torres Strait Islander young people, having appropriate information and a perception of safety in accessing treatment is necessary (Brown et al., 2016). Initial engagement characteristics of acceptance, validation, helpfulness, and flexibility in treatment approaches are essential facilitators for young men (Rice et al., 2018). At a service level, structural facilitations for help-seeking in young men include having the availability of male clinicians, strong
partnership with community agencies, and targeted messages for the importance of help-seeking for males.

Young people are also more likely to show higher rates of treatment drop-out and missed appointments (O'Brien et al., 2009). This might be due to their perceptions around the usefulness of professional help and stigma related to seeking support. The higher levels of disengagement are mainly with young people who are males, Aboriginal or Torres Strait Islanders, aged over 18 years, or living in rural areas (Seidler et al., 2020). However, a high number of those who discontinue treatment are shown to reengage in the future, and those young people may need to engage multiple times to meet their mental health needs. A study comparing the engagement of young adults (aged 18 to 25 years) to older adults (over 25 years old) in AMHS found a similar level of attendance at the initial assessment (Roche, O'Sullivan, Gunawardena, Cannon, & Lyne, 2020). However, older adults were four times more likely to have good engagement than the young adults, shown in the subsequent levels of attendance. Intervention strategies to increase engagement at an individual level have been shown to improve attendance and reduce premature treatment drop-out (H. Kim, Munson, & McKay, 2012).

Another factor relevant to engagement for young people is that they have higher rates of deterioration during therapy, and clinicians have lower rates of accurately predicting deterioration compared to adults in mental health treatment (Nelson et al., 2013; Warren et al., 2009). Specifically, deterioration rates of up to 21% have been seen in youth psychotherapy settings, compared to the 10% deterioration that is evident with adults (Murphy, Rashleigh, & Timulak, 2012; Warren et al., 2009). It is recommended that outcome measurement systems that include early warning signals are implemented in such environments. These signals would alert clinicians to deterioration and would prompt them to
explore factors for engagement with the client, such as therapeutic alliance and treatment expectations (Lutz, Schwartz, Martín Gómez Penedo, Boyle, & Deisenhofer, 2020).

**2.2.2 Youth-focused services.** Traditional mental health services are organised around a demarcation between CAMHS and a major transition to AMHS at the age of 18 years. Young people that experience this transition report uncertainty and nervousness, a lack of knowledge and preparation, and recommend that services should be more flexible and supportive around this transition period (Cleverley, Lenters, et al., 2020). Guidelines targeting this transition period propose several principles, including consultation with young people and carers, developmentally appropriate support, person-centred care, and shared responsibility from both adolescent and adult services (Willis & McDonagh, 2018). It is now well understood that this service dichotomy is particularly inappropriate for mental health services, as it undermines continuity of care at a time of greatest need and heightened vulnerability (McGorry et al., 2013).

Mental health reform in Australia has reoriented mental health services to focus on early intervention for young people aged 12 to 25 years, and this approach has been gaining momentum in other countries (Hetrick et al., 2017). *headspace* is Australia’s Youth Mental Health Foundation, which was initiated in 2006 to provide evidence-based early intervention mental health services to young people aged 12 to 25 years (Rickwood et al., 2018). *headspace* centres have been implemented progressively across Australia scaling up to a national network of over 110 centres in 2020. Internationally, several similar services have been created, including *Jigsaw* in Ireland, *Maisons des Adolescents* in France, *Youth One Stop Shops* in New Zealand, and *Foundry* in Canada (Hetrick et al., 2017). Early research has shown these initiatives have been engaging for young people, including both those with sub-threshold symptoms and those with established mental health disorders (Rickwood, Telford, Parker, Tanti, & McGorry, 2014). Moreover, youth-specific services are able to target hard to
reach groups, such as young people who identify as Aboriginal or Torres Strait Islander, those from regional and rural areas, and those disengaged from education and employment.

These youth-focused services have been created with a focus on factors such as youth participation, SDM, easy early access, evidence-informed approaches, and partnerships and collaboration (McGorry et al., 2013). A youth-friendly setting is a common feature across services, with youth involvement in choosing décor and designs, including couches, bean bag chairs, colourful walls, and creative artworks (Settipani et al., 2019). Most of these services provide open waiting areas to promote a feeling of space and safety. There is an emphasis on informal, non-clinical spaces and most services are centrally located or close to public transportation. Other essential features of these services include being partly staffed by young people, having timely appointments, and being free or low cost. There is an emphasis on maintaining confidentiality and privacy, integration of a range of services, and delivery of safe and appropriate interventions (Hetrick et al., 2017). This has led to the development of tools and supports aimed at engaging young people, such as age-appropriate psychosocial and mental health assessments (Bradford & Rickwood, 2015a). New methods of delivering mental health interventions to young people have also emerged, which include online and smartphone applications for counselling, self-help, assessment, and support groups (Dowling & Rickwood, 2013).

Evaluations of youth mental health services show promising results and there is increasing innovation in the area, targeting evidence-based practice with young people and novel tools for engagement (Mei et al., 2020). Service satisfaction, in particular session satisfaction, expectations, and organisation satisfaction, has been shown to be generally high in youth-focused services (Rickwood et al., 2017). Furthermore, client satisfaction has been shown to increase with ongoing face to face sessions. Outcome-based research on headspace services shows that after treatment 60% of clients have significantly lower levels of
psychological distress or significantly higher levels of social and occupational functioning (Rickwood, Mazzer, Telford, et al., 2015). A follow-up study exploring *headspace* services showed that clients experienced improvements in psychological distress and quality of life while engaged in services and these gains were maintained after exiting the service, even up to two years later (headspace Research and Evaluation, 2019). Similarly, a review of youth mental health services worldwide found eight studies that showed 52% to 68% of young people experienced reductions in symptoms after treatment (Hetrick et al., 2017).

Concerningly, service evaluations have been limited by the lack of appropriate outcome measures for youth across the adolescent and early adult life stages (Hilferty et al., 2015).

### 2.3 Mental Health Outcome Measurement

#### 2.3.1 Broad mental health outcome measures. An outcome measure in mental health care can be defined as a tool used to evaluate the effect on a person’s mental health as a result of health care intervention, plus any additional extra-therapeutic influences (M. Slade, 2002). Specifically, outcome measures are quantitative indicators used at two or more points in time: baseline, during treatment, discharge, or follow-ups (AMHOCN & CMHA, 2013). Routine outcome measurement, whereby the same outcome measure is used frequently at several time points, has been adopted in CAMHS and AMHS across Australia, New Zealand, Denmark, United Kingdom, and Norway (Burgess, Pirkis, & Coombs, 2015; Roe, Drake, & Slade, 2015). This push has been driven by an increasing emphasis on quality assurance and effectiveness of services. Routine outcome measurement reported at a government level enables decision making around the funding of services, particularly where health resources are limited and need to be distributed appropriately to achieve the best outcomes (McKay, Coombs, & Pirkis, 2012).
Outcome measures are increasingly designed to measure broad mental health status rather than assess symptoms associated with the diagnosis of specific psychiatric disorders (Bentley, Hartley, & Bucci, 2019). Specific measures may be required for diagnosis, but are not helpful when making comparisons between cases and services where differences in case-mix exist (Deighton et al., 2014). Using specific diagnostic measures also mean clinicians need to isolate a particular presenting problem at baseline to assess subsequent change. This presents challenges when clients have comorbid mental health issues, or their presenting problems change throughout treatment (Bearman & Weisz, 2015). In contrast, being generically relevant to a broad range of mental health presentations enables the measure to cater for clients with no disorders, such as those accessing prevention programs, through to those with severe disorders, such as public mental health clients (Bentley et al., 2019).

### 2.3.2 Domains of mental health outcome measures

Mental health is a broad construct that can be measured via many different domains (AMHOCN & CMHA, 2013). To be clinically useful, outcome measures need to be meaningful to clients and relevant to the areas in which they have treatment goals. Research with mental health service consumers reveals that many measures are not particularly relevant to their situation and do not capture personally meaningful outcomes (Crawford et al., 2011). Determining an outcome measure that is applicable in both clinical work and service evaluation is challenging (Wolpert et al., 2014). Several mental health outcome measurement domains have been proposed, which include measures that cover recovery, cognitive and emotional symptoms, functioning, and quality of life (AMHOCN & CMHA, 2013). Each domain has been recognised as providing a meaningful aspect of a client’s mental health status, but may vary in value for clinical use, service evaluation, and epidemiological studies (Wolpert, Cheng, & Deighton, 2015).

Traditionally, mental health recovery referred to attaining a cure from mental illness and was generally applied to adults with long-term severe mental illness (Andresen, Caputi,
& Oades, 2010). A contemporary definition of recovery is that it is a personal, self-defined process of change and growth towards wellbeing (Collier, 2010). Several psychological processes have been introduced in the conceptual framework of recovery, including connectedness, hope and optimism about the future, identity, meaning and purpose in life, and empowerment (Leamy, Bird, Le Boutillier, Williams, & Slade, 2011). A large number of recovery measures exist and comprehensive reviews have identified the capacity for some of these measures to be used routinely, including the Recovery Assessment Scale, Recovery Star, Recovery Process Inventory, Illness Management and Recovery Scales, and Stages of Recovery Instrument (AMHOCN & CMHA, 2013; Burgess, Pirkis, Coombs, & Rosen, 2011). However, there are concerns that consumer-defined recovery shows little association with clinical measures and there is limited evidence of the routine use of recovery measures in clinical practice (Andresen et al., 2010; Burgess et al., 2011).

Cognitive performance and emotional experience comprise a broad domain with a large number of measures (AMHOCN & CMHA, 2013). These include measures of psychological distress, anxiety and depression, general and specific measures of symptomatology, and measures of cognitive capacity. Some popular measures in this domain include the Depression, Anxiety and Stress Scales-21 (DASS-21) (Lovibond & Lovibond, 1995), General Health Questionnaire-12 (GHQ-12) (D. P. Goldberg, 1972), and Kessler Psychological Distress Scale 10 (K10) (Kessler et al., 2002). The K10 is included in the Australian National Outcomes and Casemix Collection, which is a suite of measures that have been thoroughly tested and can be used across many settings (National Mental Health Information Development Expert Advisory Panel, 2013).

Functioning measures usually focus on two aspects: activities of daily living and interpersonal relationships. Measures focussing on activities of daily life cover both simple and complex developmentally appropriate living activities. Interpersonal relationship
measures consider the developmentally appropriate quantity and quality of personal relationships, which can include partners, families, and friends (AMHOCN & CMHA, 2013). An example of a measure of functioning is the Global Assessment of Functioning, which was included in the revised third and fourth editions of the Diagnostic and Statistical Manual (Startup, Jackson, & Bendix, 2002). It measures overall psychiatric disturbance, integrating three different dimensions of functioning within a single score: psychological, social, and occupational. The World Health Organization Disability Assessment Schedule 2.0 was chosen in favour of the Global Assessment of Functioning for the Diagnostic and Statistical Manual Version 5, due to limitations of the Global Assessment of Functioning as a single-item measure, which include its lack of conceptual clarity, questionable psychometric properties, and the training required for its administration (Gold, 2014).

Quality of life measures focus on general life satisfaction, wellbeing, and personhood (AMHOCN & CMHA, 2013). A review examining quality of life measures identified the most common and essential domains, which include employment, health, leisure, living situation, and relationships (Connell, O’Cathain, & Brazier, 2014). Some typical quality of life measures are the EuroQOL-5Dimensions (Herdman et al., 2011) and the 36-Item Short-Form Health Survey (Ware & Sherbourne, 1992). However, there have been some concerns raised around the use of these generic preference-based quality of life measures, primarily due to their overemphasis on physical health symptoms and lack of content validity (Connell et al., 2014). A synthesis of qualitative research in this area identified six domains – wellbeing, control, autonomy, self-perception, belonging, activity, and hope – and recommended that these themes be incorporated into the development of future quality of life measures, to provide a more holistic mental health measure that covers the full spectrum of a person’s life (Connell, Brazier, O’Cathain, Lloyd-Jones, & Paisley, 2012).
Lastly, measures have been created to cover aspects across multiple domains, which can sometimes also include areas around service satisfaction (AMHOCN & CMHA, 2013). A suite of multidimensional measures is the HoNOS family of measures, which include the HoNOS for adults and the HoNOSCA for children and young people under the age of 18 years (Pirkis et al., 2005). Both measures cover domains of behaviour, impairment, symptoms, and social functioning. These measures are primarily clinician-rated and are also part of the Australian National Outcomes and Casemix Collection (National Mental Health Information Development Expert Advisory Panel, 2013).

2.3.3 Session-by-session outcome measures. There is expanding implementation of session-by-session measures, whereby clients complete outcome measures at each session that can be used in MFS (Hall et al., 2015). MFS allows clinicians to quantitatively track clients’ progress, which allows for data-informed decisions utilising measurement information (Miller, Duncan, Sorrell, & Brown, 2005). Measures administered at every session can also allow clinicians to discuss treatment progress with a client routinely. Clinicians and clients have identified that these measures need to be brief, sensitive to change, and meaningful to both clients and clinicians (Hall et al., 2014). Notably, measures administered electronically using computer or tablet devices are engaging for young people (Hall et al., 2015). The inclusion of timely feedback using computer-enabled visual graphs to display outcomes are useful to both young people and clinicians. Several measures have been designed for session-by-session use, including the Outcome Rating Scale (ORS) (Miller et al., 2005), Youth Outcome Questionnaire (Y-OQ) (Ridge, Warren, Burlingame, Wells, & Tumblin, 2009), Clinical Outcomes in Routine Evaluation-10 (CORE-10) (Barkham et al., 2013), Session-by-Session Strengths and Difficulties Questionnaire (SxS-SDQ) (Hall et al., 2015), and Symptoms and Functioning Severity Scale (Athay, Riemer, & Bickman, 2012).
One of the main requirements of session-by-session measures is the brief length, while providing an efficient assessment that covers a range of meaningful mental health areas. While longer questionnaires can increase the internal consistency of a measure, they can also increase fatigue in respondents, thereby reducing the accuracy of responses (Smits & Finkelman, 2014). Existing lengthy self-report measures that are required to be completed routinely can be burdensome and can reduce a client’s willingness to complete the measure. Furthermore, clinicians may be less likely to utilise data from longer measures as there is increased difficulty interpreting such information (Bentley et al., 2019). However, longer measures provide more detailed assessment information to clinicians, which can help better inform treatment planning and goal setting. Longer measures have traditionally been used in monitoring and evaluation of programs, and it is suggested that short repeated measures can help monitor treatment progress, whereas longer measures are better suited to assess the impact of clinical programs (Haroz et al., 2020).

There is increasing research to show that many measures can be shortened to be more efficient, while still providing the same ability to track change (Haroz et al., 2020). Some methods that have targeted removing unnecessary items in a measure include the use of item response theory and computerised adaptive testing (Haroz et al., 2020; Smits & Finkelman, 2014). Item response theory analysis retains the most critical items, which allows for shortened measures that show comparable psychometric properties to the standard longer measures. Furthermore, when analysing treatment outcomes using the same data, the results of a condensed measure can show comparable effect sizes to the longer standard scale, showing the same ability to capture change with less burden on clients. Such studies show how shortened measures can be used for clinical purposes and still be appropriate for program evaluation and research. For example, a study conducted in Thailand examined a 17-item depression scale that was shortened to seven items, and the 30-item Posttraumatic Stress
Scale that was reduced to 10 items, using item response theory, and these were useful both clinically and for evaluation purposes (Haroz et al., 2020).

Brief measures with very few items, such as the ORS, appeal to services in which regular outcome monitoring is required and for clinicians to use in MFS. However, due to their brief nature, such measures may not provide clinicians with an in-depth understanding of a person’s difficulties to inform treatment goals (Bentley et al., 2019). For example, the two and nine item Patient Health Questionnaires (PHQ-2 and PHQ-9) have been compared on ratings of sensitivity and specificity (Arroll et al., 2010). When clinicians used a threshold score of two or higher, the PHQ-2 had a rating of 0.86 sensitivity of predicting Major Depression Disorder, which was comparable to the PHQ-9. However, the PHQ-2 only had a modest specificity rating of 0.78, which was lower than the PHQ-9. It was recommended that if a client scores two or higher on the PHQ-2, they continue to complete the PHQ-9. This would mean still identifying clients at risk of depression, while reducing how many questions those at less risk would have to complete. These results point to the need to find a balance with the number of items needed in a measure and what is considered meaningful.

2.4 Mental Health Outcome Measures for Young People.

2.4.1 Development of outcome measures for young people. There is a long history of outcome measures for AMHS and CAMHS, including both community-based and inpatient settings. In Australia, a comprehensive report on outcome measurement in community settings identified 136 measures, of which 31 were deemed most appropriate and relevant for children and adolescents, adults or older persons (AMHOCN & CMHA, 2013). These measures incorporate both client and clinician respondents, and parent-report measures are available for children and adolescents. While there are numerous measures available for young people, outcome measures specific to the youth transition period of adolescence and
young adulthood are urgently required for continuity of care (Brann, Alexander, & Coombs, 2012).

Mental health outcome measures must be psychometrically tested and meaningful to the developmental group they are targeted towards (Happell, 2008). In light of this, there are now measures that have been tested for use across the 12 to 25 year age range. The Strengths and Difficulties Questionnaire (SDQ), created for children and adolescents, has recently had preliminary testing with a young adult sample aged 18 to 25 years, and results showed similar psychometric properties to those for an adolescent sample (Brann, Lethbridge, & Mildred, 2018). However, some critical distinctions were found: young adults tended to rate peer problems and the impact of difficulties on their lives higher than the adolescent sample. At the same time, parents reported young adults having significantly more emotional problems on the SDQ than parents of adolescents. This is consistent with existing research reporting higher levels of psychological distress for young adults compared to adolescents (Van Droogenbroeck, Spruyt, & Keppens, 2018). These trends are also seen in youth mental health settings in Australia when using the K10 (Rickwood et al., 2014) and Ireland when using the YP-CORE and CORE-10 (O'Reilly, Peiper, O'Keeffe, Illback, & Clayton, 2016).

Similarly, another study compared the HoNOSCA (under 18 years) and HoNOS (18+ years) when used with young adults, to determine which measure would be more suitable (Brann et al., 2012). Results showed that both measures were strongly correlated, however, the HoNOSCA reported greater severity of symptoms in the total score. An additional concern was the “problems with family life and relationships” item in the HoNOSCA was not correlated to the “problems with relationships” item in the HoNOS, and there are no other items relating to family relationships in the HoNOS (Pirkis et al., 2005). Therefore, the HoNOS possibly underestimates the importance of family with young adults, which is a
significant limitation if implemented in youth mental health settings (Brann et al., 2012). It is suggested that the HoNOSCA be utilised with young adults to reduce the issues around transition in measures, but also as it may be more reflective of what is relevant to young adults.

Measures are progressively being created or tested for young people in other domains, such as recovery, quality of life, and functioning. For example, the Recovery Questionnaire was developed to identify the recovery-relevant processes in young people treated in specialist mental health services (John, Jeffries, Acuna-Rivera, Warren, & Simonds, 2015). The measure was shown to be psychometrically sound for young people aged 10 to 18 years, however, the factor structure was not tested due to the small sample size in the validation study. It has been shown to have generally good face validity and acceptability by young people, however, similar to adult recovery measures there is mixed evidence for its association with clinical measures and sensitivity to change (Bentley et al., 2019).

The Columbia Impairment Scale-Youth Version, originally designed to be used for children and adolescents up to the age of 17 years, has now been validated for young people aged 15 to 24 years, targeting functional impairment (Cleverley, Brennenstuhl, & Henderson, 2020). The results showed good internal consistency and a unidimensional factor structure. However, due to the small sample size a confirmatory factor analysis and measurement invariance across sub-groups, such as age and gender, could not be tested. Furthermore, this measure was only tested within a youth addiction service, and further testing is needed for it to be clinically useful for other settings.

The World Health Organization Quality of Life-BREF, which was designed for an adult population and covering quality of life areas of physical, psychological, social, and environment, has been tested with adolescents and young adults (Skevington, Dehner,
Similar psychometric properties were seen for both groups, however, some differences were observed. Adolescents perceived greater opportunities to access information and skills, and perceived their home environment as better compared to young adults. Overall, it was recommended that adjustments to the World Health Organization Quality of Life-BREF were required for it to be suitable for young people, with a reduced emphasis on physical health and increased focus on social domains.

As can be seen from the above review of studies that have examined mental health measures used with young people, there is the need for meaningful measures that are psychometrically tested for the entire 12 to 25 year age range, with large samples that are sensitive to change and can be utilised with a broad youth population.

2.4.2 Session-by-session outcome measures for young people. Session-by-session measures for youth are also expanding for use in MFS. The ORS, initially tested for an adult population, has also been tested with young people aged 13 years and above with the intent to be used in MFS (Duncan, Sparks, Miller, Bohanske, & Claud, 2006). The ORS was shown to be generally suitable for young people and provided a developmentally distinct cut-off score for adolescents. The SxS-SDQ, the adaption of the SDQ for sessional use, consists of four questions related to difficulties, distress, social impairment, and progress (Hall et al., 2015). The SxS-SDQ was implemented within an electronic system that generated instant reports for clinicians, including graphs of previous scores. Results showed better completion rates by young people of the SxS-SDQ, compared to other measures, and this was mainly related to the electronic system allowing ease of use (Hall et al., 2015).

The Young Person-CORE (YP-CORE) and CORE-10, both 10-item measures, have been designed for sessional use (Barkham et al., 2013; Twigg et al., 2009). When tested in a youth mental health service across the 12 to 25 year age range, the YP-CORE was used for
those aged 16 years and under, and the CORE-10 was used for those aged 17 years and older (O'Reilly et al., 2016). The results showed that the measures were psychometrically sound for their respective age groups. Both measures had single-factor structures, and multigroup confirmatory factor analyses indicated structural equivalence across gender and age groups.

The design of session-by-session measures for young people is growing, however, there is still a lack of these measures designed specifically for young people aged across the 12 to 25 year age span. The creation of developmentally appropriate session-by-session measures that can be administered electronically and used in MFS is crucial (Donald et al., 2018). This would allow clinicians and organisations to continue to track change with clients transitioning from adolescence into young adulthood, which is shown to be a vulnerable period for young people and, traditionally, has shown to be a disruptive time in mental health service provision with movement from CAMHS to AMHS (Cleverley, Lenter, et al., 2020; Donald, Rickwood, & Carey, 2014).

### 2.4.3 Electronic administration of outcome measures.

Increasingly, mental health outcome measures are being adapted to be administered electronically. Equivalence between paper and electronic versions of measures has been explored through comparing mean scores by correlations and t-tests, and there is evidence to show general similarities with high correlations and small differences (van Ballegooijen, Riper, Cuijpers, van Oppen, & Smit, 2016). In youth literature, there are generally higher ratings of user satisfaction with electronic versions of outcome measures compared to paper versions (Truman et al., 2003). Specifically, electronically administered measures can hold young people’s attention better, meaning better response rates, reduced completion time, and greater measurement validity.

Electronic administration is particularly useful for clients who might find it challenging to communicate targeted problems to clinicians and those who dislike filling in
forms or find them difficult to complete (Truman et al., 2003). This becomes particularly important when completing measures every session. For example, the electronic SxS-SDQ was compared to the use of the paper version of the SDQ in routine practice (Hall et al., 2015). The electronic SxS-SDQ showed higher completion rates by clients. Qualitatively, the electronic version of the SxS-SDQ was described as “fun” and “cool” by young people and was more desirable than using a paper copy of the measure (Hall et al., 2014).

There are, however, mixed views from clinicians about electronic administration of outcome measures, with some concerns with the use of technology negatively impacting on the therapeutic alliance (Youn, Kraus, & Castonguay, 2012). Others have expressed concerns around the additional effort required to use such systems and, specifically, difficulties around learning new electronic systems (Kaiser, Schmutzhart, & Laireiter, 2018). When working privately or for a smaller organisation, there can be concerns about the financial cost to utilise electronic measurement systems. While electronic systems can overcome some barriers for the use of outcome measures, there are also concerns around the system crashing (Hall et al., 2014). One of the main benefits of electronic tools identified by clinicians is the ability to efficiently gather input from the client to help treatment planning (Bradford, Rickwood, & Boer, 2014; Kaiser et al., 2018). For clinicians, electronic administration yields fewer coding errors and automatic scoring systems can provide instant graphs to compare against past progress (Truman et al., 2003).

Self-administered electronic outcome measures can provide young people with increased control over the help-seeking and treatment process, by allowing them to identify to the clinician what areas they are finding difficult, and also what areas are going well (Bradford & Rickwood, 2015b). A review exploring self-disclosure between online and offline systems found higher rates of self-disclosure of personal issues when using an online system (Nguyen, Bin, & Campbell, 2012). An evaluation of an electronic psychosocial
assessment tool used with young people was found to be widely accepted by both young people and clinicians (Bradford & Rickwood, 2015a). Particularly noteworthy was that the automated tool resulted in greater reporting of more sensitive issues. Reports of alcohol or substance use, sexual orientation when identifying as non-heterosexual, sexual history, past self-harming behaviours, and risk-taking behaviours were disclosed at rates 2.78 to 10.38 times higher using the electronic system, compared to the standard method of face to face assessment. This electronic application was rated by young people as highly efficient for use in session while having no negative impact on ratings of the therapeutic alliance.

Similarly, a comparison of the K10 measure being used with adults through a face to face interview and an online survey found that the online survey produced higher scores of psychological distress than face to face interviews (Klein, Tyler-Parker, & Bastian, 2020). Those aged 18 to 34 years were also more likely to have higher K10 scores than their older counterparts, and the discrepancy of scores between face to face interviews and online surveys was greater for the younger aged participants. It has been suggested that electronic measures can elicit increased comfort, decreased inhibitions, and more truthful responses, particularly in younger clients and those more prone to higher psychological distress. An explanation for this is that young people are more experienced with the use of electronic technology and hence more comfortable in revealing personal information in this format (Bradford & Rickwood, 2015a).

2.4.4 Developmentally specific domains of outcome measures for young people.

As previously discussed, essential aspects of mental health outcome measures are that they be meaningful and capture areas of importance for the targeted group. Moses and Claypool (2018) completed semi-structured interviews with adolescents six months following a first-time, voluntary, and brief psychiatric hospitalisation. They aimed to capture the adolescents’ description of what had changed in their wellbeing over time and revealed four major themes:
symptoms, relational, coping, and a sense of purpose. Symptoms for young people related to emotional and mental wellbeing and these items are highly represented in many gold-standard youth measures, including the HoNOSCA, SDQ, and YP-CORE (Brann et al., 2018; Pirkis et al., 2005; Twigg et al., 2009).

Interpersonal relationships are essential for this age group as young people rely on peer relationships, even while still living with family for more extended periods (Spear, 2000). As previously mentioned, when testing the HoNOSCA and SDQ to be used with young adults, interpersonal relationships were specifically identified as important for the whole age range of 12 to 25 years (Brann et al., 2012; Brann et al., 2018). Even when young people have positive relationships, they still report negative changes in relationships after a mental health diagnosis or treatment (Woodgate et al., 2017). Generally, the quality of relationships are not prominently featured in post-treatment measures of effectiveness, but for young people, it may serve as a benchmark for their mental health status (Moses & Claypool, 2018).

Coping measures are also not typically used for gauging the effectiveness of treatment, but again they have been identified as important for mental health in young people (Moses & Claypool, 2018). Coping is a broad concept that relates to cognitive and behavioural efforts to manage mental health. Coping strategies with adolescents and young adults can include fostering abilities, cognitive restructuring, and engaging in pleasurable activities (Summerhurst, Wamnes, Wrath, & Osuch, 2017; Woodgate et al., 2017). A sense of purpose relates to engagement with school, work, and being future-orientated (Moses & Claypool, 2018). These day-to-day activities have also been emphasised in existing youth measures, such as the HoNOSCA and SDQ (Brann et al., 2018; Pirkis et al., 2005). Such domains of mental health should be incorporated into outcome measures for young people.
While being a distinct developmental group, there are also gender differences across the 12 to 25 year age range for mental health outcomes. In community samples, gender differences are seen across psychological distress, anxiety, and depression, with females generally reporting significantly higher scores than males (Van Droogenbroeck et al., 2018). Furthermore, when exploring measures of functioning and quality of life, young females have been shown to have lower levels on both domains of mental health, compared to young males (Cleverley, Brennenstuhl, et al., 2020; Rissanen, Lindberg, Marttunen, Sintonen, & Roine, 2019).

2.5 Tracking Change Using Outcome Measures

2.5.1 Paired t-test and effect sizes. A vital issue has been finding agreement on what constitutes clinically meaningful change with various quantitative indices being proposed (Wolpert, Görzig, et al., 2015). When measuring effectiveness, fundamentally, outcome measures must be sensitive to change and be able to clearly convey the magnitude of change achieved. The most basic method of exploring this is using analysis of statistical significance, such as a paired t-test, to identify changes in group mean scores on an outcome measure before and after treatment (Evans, Margison, & Barkham, 1998). However, statistical significance in such analyses does not always equate to clinical significance, and even if a change is statistically significant it may not be meaningful to the client (Burgess, Pirkis, & Coombs, 2009; Wise, 2004). These statistical analyses are sensitive to sample size with small samples providing less confidence in results (Wolpert, Görzig, et al., 2015). In contrast, large samples can produce statistical significance even with small pre-post treatment differences. Additionally, calculating the differences in scores to a standardised effect size provides further information, however, it still does not always translate to meaningful change (J. Cohen, 1988). Use of statistical significance and effect sizes can be more useful when comparing with a control group (Wolpert, Görzig, et al., 2015).
2.5.2 Reliable change indexes and clinically significant change indexes. An increasingly used technique for determining an appropriate level of change in outcome measure scores is using RCI. Reliable change takes into account the reliability of the measurement instrument and ensures that change is not due to measurement error (Jacobson & Truax, 1991). Less reliable outcome measures require a more substantial difference in change scores to achieve reliable change, and this method has been proposed to provide a more accurate standard of meaningful change (Wise, 2004). The Jacobson and Truax (1991) method for calculating reliable change takes the change in a client’s score and divides it by the standard error of the difference between the pre-post measurement scores. A reliable change that is greater than 1.96 ($p < .05$) means that the change in score is unlikely to occur without actual change and that the difference is not just due to the fluctuations of an unreliable outcome measure. From this, RCI can be calculated for a specific mental health outcome measure for a particular population, to get an indicator of how much change is needed on that measure for reliable change to occur.

The RCI for the ORS provided by the original study with an American sample is a change of five points (Miller & Duncan, 2004). However, when the ORS was tested with a Dutch sample, the RCI had a higher index of nine (Janse, Boezen-Hilberdink, van Dijk, Verbraak, & Hutschemaekers, 2014). Similarly, in another study targeting a Dutch sample, specifically from a Jewish mental health service, the RCI for the ORS was eight (Hafkenscheid, Duncan, & Miller, 2010). Therefore, a greater change in the ORS score is needed for the Dutch population compared to the American population, for the change to be deemed reliable. This has important implications for clinician and client decision making when the RCI is used in MFS, as cultural factors can affect the reliability of a measure (Janse et al., 2014).
Further, estimating clinical significance, which is distinct from statistical significance, has been recommended in mental health contexts. CSI can show if a client starting treatment in the dysfunctional (clinical) population, completes treatment no longer in this population (Jacobson, Follette, & Revenstorf, 1984). A clinically significant change is operationalised as change on a client’s outcome measure score showing that they are statistically more likely to be drawn from the functional distribution, having moved out of the dysfunctional distribution during treatment (Jacobson & Truax, 1991). When the dysfunctional and functional populations are identified, CSI can be calculated by finding the value where the two populations intersect (Wise, 2004). This technique recommends that comparison populations and norms are used, however, in many studies a normative sample cannot be identified. Similar to the RCI, CSI are estimated based on specific clinical and normative populations.

An examination of clinically significant change using the ORS in a Dutch sample found that the CSI was a score of 24, one point lower than the American cut-off score in the original ORS validation study (Janse et al., 2014). This was likely due to the ORS total score in the clinical group for the Dutch study \((M = 17.0, SD = 7.2)\) being lower than the clinical group reported by Miller, Duncan, Brown, Sparks, and Claud (2003) \((M = 19.6, SD = 8.7)\). This meant the clinical group from the Dutch study would need to reach a lower score of 24 compared to the American cut-off of 25 to reach clinically significant change. Additionally, for young people between the ages of 13 to 17 years from the USA, the CSI on the ORS is 28 (Duncan et al., 2006).

A problem with using the CSI is that it cannot be calculated for clients who are not in the clinical population when they commence treatment. This is a particular issue for early intervention mental health services, where some clients may present initially in the functional, non-clinical range (Rickwood, Mazzer, Telford, et al., 2015). This means they cannot achieve clinically significant improvement, even though they may substantially
improve. Furthermore, clients initially presenting with outcome measurement scores close to the CSI can achieve clinically significant improvement more easily than clients presenting with scores further from these indexes, and these clients may have achieved very little substantive change (Wolpert, Görzig, et al., 2015).

It is recommended that RCI and CSI are utilised together to assure clinically significant change is also reliable. Therefore, for treatment outcomes to be considered reliable and clinically significant, two criteria need to be met: change must be shown to be statistically reliable (RCI) and the client must move from the dysfunctional to the functional distribution (CSI) (Wise, 2004). Change can then be categorised into four stages: deterioration - when a client has reliably worsened; unchanged - when no reliable change has occurred; improvement - when a client has made a reliable positive change but remains in the dysfunctional population; and recovered - when a client reliably improves and moves into the functional population (Jacobson & Truax, 1991).

Paired t-tests, effect sizes, RCI, and CSI all provide different types of information of use to organisations (Wise, 2004). For example, a study exploring the outcomes in youth mental health services using the K10, showed 36.1% of clients reaching improvement with an effect size of 0.5 or greater, 26.2% reaching reliable improvement, and 21.2% reaching clinically significant improvement (Rickwood, Mazzer, Telford, et al., 2015). Only those initially presenting in the clinical range could be tested for clinically significant change, however, and therefore a smaller sample was used. Another example is a study from the United Kingdom that used the CORE-OM to assess effectiveness across mental health services (Stiles, Barkham, & Wheeler, 2015). The inclusion criteria consisted of clients who initially presented in the clinical range, those presenting up to 40 sessions, and those who had planned endings. Results showed that the pre-post treatment changes had a large effect size of
1.89, 79.9% showed a reliable improvement, and 60% were deemed recovered (i.e. showed reliable and clinically significant improvement).

While more measures have been psychometrically tested to be used with young people, there has also been a further exploration in providing norms and CSI based on age groups and gender across different settings. Norms for the SDQ, initially provided for the United Kingdom, have now been calculated for the Australian population, and significant age and gender differences were seen across the 7-17 year age range (Mellor, 2005). Males had significantly higher self-reported difficulties scores on conduct problems and hyperactivity scales, while females had higher self-reported difficulties scores on emotional symptoms and prosocial behaviour scales. A study exploring the differences across pre-adolescence, early adolescence, and late adolescence on the SDQ showed support for the five-factor structure of the SDQ in all three age groups in a Norwegian sample (Van Roy, Veenstra, & Clench-Aas, 2008). However, results presented unequal factor loadings across age groups with loadings relatively similar for early and late adolescents, and generally lower for the pre-adolescent group. These results indicate the need for age and gender-specific norms and CSI, because of age and gender differences in mental health outcomes for this age group.

2.5.3 Growth curves (expected change trajectories). A criticism of CSI is that it can be an overly stringent measure of change, being based on diagnostic cut-offs. In naturalistic mental health settings, some clients may not be able to reach this threshold because they initially present in the functional population range (Wise, 2004). Other methods of monitoring change have been recommended, specifically, the use of growth curve modelling, which shows expected rates of change (Donald & Carey, 2017).

Growth curve modelling estimates a mean starting point (intercept) and average rates of change (slope) of the pooled sample trajectory, that is, within-person expected change
patterns (Singer & Willett, 2003). Also known as expected change trajectories or expected treatment response curves, growth curves provide information about the average amount of change that occurs, how fast changes are made, and the pattern of change. This provides an additional level of information to organisations in exploring service effectiveness, while for academics it has useful implications for clinical research and theory (Warren, Nelson, Mondragon, Baldwin, & Burlingame, 2010). The method is particularly helpful in exploring client change in naturalistic therapy settings, as it can deal with data that are time-unstructured and unbalanced. It requires data to be collected frequently and the creation of session-by-session measures has allowed more research to be conducted that compares rates of change across settings (Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009).

Patterns of growth curve modelling can be linear or non-linear. The linear slope indicates that the rate of change remains constant across time, and the slope is a straight line (Singer & Willett, 2003). The higher-order polynomial trends show that rate of change amongst clients might not be the same over time. A quadratic (second-order polynomial) slope has a change trajectory that has no constant common slope while having a single stationary point in the form of a peak or trough. This pattern of change has a parabola shape with an initial shift in measurement scores occurring quickly at first, then tapering off over time (Shek & Ma, 2011). A cubic slope has two stationary points, with one peak and one trough. It has a steep increase during the early stages of therapy, subsequent flattening out, and then a slight increase again later in treatment.

A cubic pattern of change has been observed in numerous psychotherapy settings (Baldwin et al., 2009; Falkenström, Josefsson, Berggren, & Holmqvist, 2016). Interestingly, there is evidence that gains, particularly sudden gains, occur earlier in treatment for younger clients. For example, Gaynor et al. (2003) found that in children and adolescents, 85% of sudden gains occurred before session five and all sudden gains by the 10th session. In
contrast, for an adult sample, only half of the sudden gains had occurred by session five (Tang, Luborsky, & Andrusyna, 2002). Furthermore, clients who experience early changes in therapy have been shown to be 21 times more likely to experience reliable change and 19 times more likely to reach clinically significant change (Erekson, Clayson, Park, & Tass, 2020).

Two main dose-response models attempt to explain the cubic pattern seen in psychotherapy settings. Firstly, the dose-effect model (Howard, Kopta, Krause, & Orlinsky, 1986) is based on a medical framework, whereby sessions in psychotherapy are compared to doses of medication. Similar to medication, it proposes that increasing the number of sessions will provide clients with more of the active ingredients from psychotherapy. This model also suggests that this rate of change is negatively accelerating with clients initially improving but benefits of therapy reducing in later sessions. Another theory for this pattern of change is the “good enough level” model, whereby clients are said to change at different rates (Barkham et al., 2006). It is theorised that those that have reached sufficient change terminate treatment and, therefore, a plateau is seen within the growth curve. In support of the “good enough level” model, a study exploring rates of change for clients in a university counselling service found that clients who participated in fewer sessions had relatively fast rates of change, whereas those who attended more sessions had slower rates of change (Baldwin et al., 2009).

Several factors have been shown to affect rates of change in youth psychotherapy settings. Previous literature exploring growth curves in such settings have utilised a time variable of weeks in treatment rather than session number, as it has provided a better model fit based on Bayesian information criterion (Warren et al., 2010). Furthermore, the literature exploring mental health treatment with young people has failed to demonstrate a significant dose-response relationship between sessions attended and treatment outcome (Bickman, Andrade, & Lambert, 2002). When exploring group differences in rates of change, a
precedent has been set for analysing change trajectories dependent on baseline severity on outcome measures, as this has been shown to account for the most variance (Finch et al., 2001; Lambert et al., 2002; Smith & Jensen-Doss, 2017). It has been shown that rates of change were faster for those who had more severe baseline measurement scores compared to those with less severe scores (Cannon, Warren, Nelson, & Burlingame, 2010).

Research has increasingly focused on detecting clients who are at risk of deterioration using early warning systems that are derived from expected change trajectories (Finch et al., 2001). An early warning is evident when a client’s score drops below an identified threshold. It is recommended that these early warning signals be derived from the proportion of clients who have been shown to deteriorate within the targeted population (Nelson et al., 2013). Young people in mental health treatment have shown deterioration rates of up to 20% (Warren et al., 2010). The provision of these expected change trajectories and early warning signals allows a measure to be utilised more thoroughly in MFS (Delgadillo et al., 2018). This is essential to supplement clinical judgement, to improve clinical practice and outcomes in youth mental health settings. Such systems are most effective in electronic formats as they can instantly calculate a client’s outcome measurement score, graph it, compare it to past scores, and compare it against these benchmarks (Fortney et al., 2017).

The range of methods for tracking change not only provide better ways of evaluating services and quality assurance, but also offer additional information to clinicians to better monitor progress and plan treatment (Jackson, Keir, Sender, & Mueller, 2017). RCI, CSI, and expected change trajectories provide statistically meaningful benchmarks that enhance the clinical utility and interpretability of a measure (Boswell, Kraus, Miller, & Lambert, 2015; Donald & Carey, 2017). Providing RCI and CSI that are age and gender-specific are critical for youth mental health when there is a substantial developmental change occurring in multiple domains (Mellor, 2005). Furthermore, by providing expected change trajectories,
clinicians can compare this average rate of change expected in a specific treatment setting against an individual client’s trajectory, to determine whether the client is within or outside expected rates of change, potentially indicating a cause for concern (Finch et al., 2001).

2.6 Measurement Feedback Systems

The development of routine outcome measures and session-by-session measures has supported the use of MFS (Bickman, 2008). There is an abundance of terms used across the literature to refer to this clinical use of outcome measures, including Measurement-Based Care (Scott & Lewis, 2015), Outcome-Informed Therapy (Duncan, Miller, & Sparks, 2011), Feedback Informed Therapy (Miller et al., 2015), and Routine Outcome Monitoring (Carlier et al., 2012). These approaches all describe the process whereby a clinician receives feedback on a client’s progress that can be used within the session and can help guide treatment planning (McKay et al., 2012). This provides clinicians with regular up-to-date snapshots of a client’s mental health status and shows any changes since past sessions. Clinicians are then able to monitor if clients are progressing or deteriorating between sessions, and adjust treatment planning accordingly (Boswell et al., 2015; Coombs, Stapley, & Pirkis, 2011). Such an approach can also allow clinicians to bring the measures into a session and provide feedback of measurement scores to clients, which can be a powerful way to promote SDM (De Jong et al., 2014).

2.6.1 Clinician blindspots. MFS are particularly helpful in addressing clinician blindspots. Clinicians have been shown to have low accuracy rates of predicting client deterioration during therapy when using their judgement alone (Hannan et al., 2005; Hatfield, McCullough, Frantz, & Krieger, 2010). Monitoring treatment has traditionally been at the level of clinician intuition, however, clinician experience has not been shown to be highly related to predictive accuracy, compared to measurement-based outcomes. This is particularly
worrying, as clinicians are reluctant to identify deterioration and tend to overestimate improvement rates (Hannan et al., 2005).

A study that examined clinicians’ therapy notes for 70 clients who had reliably deteriorated, measured using the Outcome Questionnaire-45 (OQ-45), revealed that clinicians using clinical judgment alone identified clients reliably deteriorating only 21% of the time, even though all examined cases had indicated deterioration on the measure (Hatfield et al., 2010). Similarly, clinicians in a university outpatient clinic were asked to estimate client progress based on their clinical judgment alone, and this was then compared with outcomes on the OQ-45 (Hannan et al., 2005). Clinicians were explicitly advised that there had been a stable 8% deterioration rate in the clinic in past years, yet results showed that clinicians estimated 0.01% of clients deteriorated, which was considerably lower than the 7.3% shown to deteriorate on the OQ-45.

Other research has explored clinicians’ self-rating of their clinical skills. Walfish, McAlister, O’Donnell, and Lambert (2012) examined clinicians’ ratings of their own clinical expertise and client outcomes, finding that clinicians rated their skills, on average, at the 80th percentile and that all clinicians rated themselves above the 50th percentile. Additionally, clinicians, on average, believed that 77% of their clients improved as a result of their therapeutic intervention. This is well above the one-third proportion of clients shown to improve in naturalistic mental health settings (Hansen, Lambert, & Forman, 2002). Similar to the study by Walfish et al. (2012), Z. J. Parker and Waller (2015) also found clinicians overestimating their abilities, however, the rating of their own clinical skills was slightly more conservative at the 66th percentile, which could be attributed to the study using a sample from the United Kingdom compared to an American sample. Results also showed, based on their own judgement, clinicians reported that over 80% of their clients improved or recovered, while only 4% of their clients deteriorated.
Another study exploring clinician self-ratings on their competence with cognitive therapy showed moderate agreement between self-ratings and ratings by an independent rater (Brosan, Reynolds, & Moore, 2008). However, further examination showed that clinicians overall tended to overestimate their competence, relative to expert judgements, and that this over-evaluation was significantly higher in less competent clinicians. This is consistent with the Kruger and Dunning (1999) effect, whereby those who are less competent believe they are most informed and knowledgeable, while those who are most competent underestimate their own abilities.

There has been work exploring potentially harmful therapies and this challenges the past misconception that doing something in clinical practice is better than doing nothing. Some psychological treatments can exert adverse effects on clients and Lilienfeld (2007) has published a provisional list of potentially harmful therapies for clinicians to avoid. A qualitative study that investigated clinicians’ perspectives and past experiences of adverse effects of psychological treatments found that the majority of clinicians identified the existence of negative effects in mental health treatments, and many reported having experienced deterioration of clients in their own practice (Bystedt, Rozental, Andersson, Boettcher, & Carlbring, 2014). However, there was a lack of consensus around what met the criteria for adverse experiences by clients. Clinicians were shown to have limited knowledge around the use of routine outcome measures in predicting client deterioration and reducing negative effects. Even when provided with information about the baseline deterioration rates, it seems that clinicians are often still unwilling, or unable, to identify client deterioration and are inclined to devalue such statistical data (Hannan et al., 2005).

This propensity to rate oneself highly is proposed to be a self-assessment bias, which is not only seen in mental health clinicians, but also in other aspects of people’s life and other professions. Self-assessment bias with clinicians can occur when cognitive dissonance arises,
when the clinician unintentionally reduces the discrepancy between their ideal self and treatment outcomes, in a self-serving manner (Dunning, Heath, & Suls, 2004). This type of behaviour can also be explained by the confirmation bias, which is the tendency to unconsciously give unwarranted preference to information that supports a clinician’s position, while dismissing information that is against it (Nickerson, 1998). Interestingly, professional self-doubt is proposed as constructive and beneficial for clinicians, resulting in more effective clinical practice (Nissen-Lie et al., 2017). It is argued that clinicians who are more conscious of their own limitations and uncertainties in a clinical capacity are less blinded by their competence. Clinicians who are more realistic about the likelihood of poor treatment outcomes are more vigilant in identifying clients who are deteriorating or not on track, enabling them to make appropriate treatment adjustments as needed (Macdonald & Mellor-Clark, 2015).

Nissen-Lie et al. (2017) investigated the interplay of clinician levels of professional self-doubt, coping strategies, and self-affiliation, and how they related to treatment outcomes. Self-affiliation is defined by one’s ability to nurture and accept oneself (Benjamin, 1996). Results showed that clinicians with higher levels of professional self-doubt had better client outcomes (Nissen-Lie et al., 2017). Treatment outcomes were even better when clinicians had both higher levels of professional self-doubt and higher levels of self-affiliation. At the same time, those with the lowest levels of professional self-doubt combined with the highest levels of self-affiliation showed the least amount of client progress. It was theorised that a healthy self-critical stance in clinical work is essential for allowing clinicians to navigate the complexity of their work, and the addition of positive self-affiliation can enhance this. An additional result of this study was that clinicians who used constructive coping strategies had better treatment outcomes. These involve actively dealing with problems, exercising reflexive control, seeking consultation, and problem-solving together with the client.
The literature reinforces the notion that clinicians need impartial data to alert them when a client is deteriorating or not on track. Research consistently supports the conclusion that formal methods of monitoring client treatment are valuable (Hannan et al., 2005). Given the distinctive pattern of therapeutic drop-out and deterioration, there is a recommendation to target trajectories of failed therapy and investigate the relative prevalence of each pattern in different therapy settings and client groups (Bowie, McLeod, & McLeod, 2016). This would allow clinicians to predict pivotal moments when clients start to find therapy unhelpful and make changes before this happens. Also, formal feedback through the use of outcome measurement provides a structured method for reducing the self-assessment bias seen amongst clinicians (Macdonald & Mellor-Clark, 2015). The research suggests tracking client outcomes using psychometrically tested measures administered routinely, if not every session. Efficient feedback of measurement scores should be provided to clinicians with early warning systems that signal when corrections are required during treatment (Lilienfeld, 2007). Furthermore, a structured format of being able to discuss measurement scores with clients in session might help identify factors impeding progress (Macdonald & Mellor-Clark, 2015). This can also help the clinician gain more information on the client’s perspective on the therapeutic relationship and their motivation for engaging in therapy.

2.6.2 Early warning systems in measurement feedback systems. A critical focus of MFS is the detection of clients who are at risk of deterioration using early warning systems that are derived from expected change trajectories (Crits-Christoph et al., 2015). An early warning is evident when a client’s score drops below an identified threshold. It is recommended that these early warning signals be derived from the proportion likely to deteriorate in the targeted population group, specifically up to 10% in adults and 21% in youth (Lilienfeld, 2007; Murphy et al., 2012; Warren et al., 2009). High drop-out rates are another primary concern in youth mental health settings, and drop-out is partly due to
clinician and therapeutic factors that may be responsive to feedback (de Haan, Boon, de Jong, Hoeve, & Vermeiren, 2013).

An essential aspect of early warning signals is the ability to accurately predict clients who are responding poorly to treatment or who are not on track before therapy is terminated (Boswell et al., 2015). Some studies have evaluated the efficacy of these signals of deterioration, alerting clinicians to clients who are falling into the bottom 10% to 20%, demonstrating detection accuracy rates of 85% to 100% when used with adult clients (Lambert et al., 2002). Lower detection accuracy rates of 69% to 88% are seen when early warning signals are used with children and adolescents, which has been justified by the higher proportions of treatment failure when compared to adult clients (Cannon et al., 2010; Nelson et al., 2013; Warren et al., 2009).

MFS vary in the amount of information they provide, with some just providing measurement scores and graphs, while others add expected change trajectories and early warning signals. A randomised controlled trial (RCT) conducted across eight National Health Service Trusts in England compared clinical outcomes across two groups: outcome feedback with the expected change trajectories and early warning signals, and outcome feedback without these additions (Delgadillo et al., 2018). The results found that in the not on track clients, there were greater positive changes on the PHQ-9, the Generalised Anxiety Disorder-7 (GAD-7), and the Work and Social Adjustment Scale in the group with the addition of trajectories and signals, compared to the group without these additions. Effect sizes were small: 0.23, 0.19, 0.22, respectively. Furthermore, clients who were deemed not on track had lower rates of reliable deterioration in the feedback group with the additional trajectories and signals (6%) compared to the feedback group without the additions (11%). This result was also seen with the whole sample, however, the difference was smaller, 4% deterioration in the feedback with additions group and 7% deterioration in the feedback without additions.
A study similar to, but smaller than that of Delgadillo et al. (2018) was conducted in Leeds, England. The study included 18 clinicians and 594 clients and found that the addition of expected change trajectories and early warning signals also produced higher efficiency of outcomes, which meant lower costs for treatment (Delgadillo et al., 2017). The results found an average cost saving of £97.54 per treatment when outcome feedback with additional trajectories and signals were implemented. Increased treatment efficiencies have also been seen when using MFS in other studies, however, the studies by Delgadillo et al. (2018) and Delgadillo et al. (2017) show the compounding outcome effects of expected change trajectories and early warning signals added to measurement graphs. These findings emphasise the importance of the availability of formalised mechanisms to guide clinical discussions (Krägeloh et al., 2015).

2.6.3 Shared decision making and treatment planning. SDM has been defined as an approach where the clinician and client deliberate about treatment options and preferences, and choose a pathway that is informed by the clinician and is consistent with the client’s values (Metz et al., 2015). There can be a number of different steps involved in SDM but generally this includes introducing the choices available to the client, explaining options to the client with the use of appropriate clinical tools, supporting the client in preferences and decision making about most appropriate treatment, and monitoring and reviewing treatment (Elwyn et al., 2012; Langer & Jensen-Doss, 2018). SDM is recommended for most health care settings for several reasons relating to ethical obligations and appropriate client informed care (Stiggelbout et al., 2012). It provides the client with the autonomy to make informed decisions, by allowing them to balance the costs and benefits of treatment. SDM is now widely recommended in mental health care with the premise that clinicians can provide expertise in diagnosis, etiology, prognosis, and treatment options, while the client has the knowledge of their own illness experience, social circumstances, and values (Metz et al.,
2015). Being able to merge the two perspectives leads to appropriate clinical and ethical decisions (McKay & Coombs, 2012). Furthermore, SDM in mental health care is proposed to increase involvement and engagement, improve decision making, enhance treatment adherence, and overall improve treatment outcomes (M. Slade, 2017).

A systematic review of the effects of SDM on treatment outcomes identified 11 RCTs (Joosten et al., 2008). These included five trials showing no differences between SDM and a control group, five trials reporting short-term positive outcomes, and one trial reporting long-term positive outcomes (but not short-term) of SDM. In a more recent systematic review exploring SDM with psychosis, the study identified six trials with small beneficial effects of SDM on treatment-related empowerment (Stovell, Morrison, Panayiotou, & Hutton, 2016). In another review exploring SDM across different settings, 39 studies met the inclusion criteria and overall 43% of studies found a significant positive relationship between SDM and client treatment outcomes (Shay & Lafata, 2015). When the review focused on affective-cognitive outcomes only, 54% of these studies showed positive effects of SDM, which reinforces the importance of SDM in mental health care.

Metz et al. (2015) proposed a five-step model of SDM when using MFS. The first step involves the introduction of the concept with the client, including discussing expectations about SDM, the role of the client, the goals of the client, and the use of MFS as an information source. When exploring a client’s view on using MFS, it is essential that the client is made aware of what the measure is used for and who has access to the results (Börjesson & Boström, 2020). Step two involves the clinician discussing the MFS outcomes with the client, by identifying any issues, understanding the problems, appreciating the feedback, and acting on the information (Metz et al., 2015). Step three involves objectively discussing treatment options, by presenting advantages and disadvantages. Clients using outcome measures self-report in a relational context, which sometimes means scoring better
than they feel or scoring worse than they feel (Börjesson & Boström, 2020). This emphasises the importance of discussing these measurement scores in a therapeutic setting to give meaning to the data. Steps four and five of SDM using MFS involve the client weighing up options in collaboration with the clinician, and a shared decision being made about a specific treatment plan (Metz et al., 2015). When the client can see MFS as being part of therapy, it allows them to become aware of self-ratings on their own wellbeing and provides them with a space to be heard.

One study showed that higher adherence to SDM using MFS was associated with significantly less decisional conflict experienced by clients and, in turn, clients had better treatment outcomes (Metz et al., 2019). Decisional conflict refers to the level of comfort and engagement in clinical decisions, with less decision conflict related to a better decision making process (LeBlanc, Kenny, O’Connor, & Légaré, 2009). High decisional conflict has been shown to be related to decisional delay, regret around treatment, and more complaints. Generally, decisional conflict is shown to be common among mental health care clients, with relatively high scores across the five subscales and the total score on the decisional conflict scale (Metz, Veerbeek, van der Feltz-Cornelis, de Beurs, & Beekman, 2018). Specifically, having a low internal locus of control, a personality disorder, and being male have been shown to significantly increase the level of decisional conflict.

2.7 Effectiveness of Measurement Feedback Systems

A review of MFS found 45 RCTs that used MFS focusing on the mental health of the client, however, these were not exclusively in mental health settings (Carlier et al., 2012). The review’s search period was from 1975 till 2009 and no RCTs with children or adolescents using MFS were found. Most studies in this review showed MFS producing efficient provision of diagnosis and quicker adjustments to treatment, particularly earlier in
treatment. Studies exploring communication between the clinician and client when using MFS were consistently positive, while there were mixed results in the exploration of MFS and agreement towards treatment. Similarly, some studies found positive results with the use of MFS and client satisfaction, but this was not always consistent. There was also ambiguous evidence for the cost-effectiveness of MFS. Generally, the identified studies targeted adult samples, which showed positive effects of MFS on treatment outcomes across a range of mental health settings, mostly in short-term therapy and particularly with clients where there were difficulties in treatment.

A systematic review evaluated the effectiveness of using OQ-45 to measure treatment outcomes in MFS (Shimokawa et al., 2010). Six studies were identified that use this system to monitor therapeutic progress and had early alert systems implemented to identify clients at risk of treatment failure. Overall, the effects of MFS on clients who were identified as being not on track were more substantial than those for clients who were showing the expected change. For clients not on track in the treatment as usual (TAU) group, 20.1% deteriorated and 22.3% had clinically significant improvement. The not on track group who had measurement feedback provided to clinicians only showed 13.6% deterioration and 30.9% improvement, while the not on track group with measurement feedback to both clinician and client showed 15.8% deterioration and 38.7% improvement. Lastly, the not on track group with measurement feedback and the use of clinical support tools showed 11.3% deterioration and 37.6% improvement. These results indicate that all feedback interventions were effective in reducing deterioration rates and increasing improvement rates among not on track clients.

The Partners for Change Outcome Management System (PCOMS) is one of the most widely known MFS, utilising the ultra-brief ORS measure that covers four items: individual, interpersonal, social, and overall wellbeing, and the session rating scale that covers four dimensions of therapeutic alliance: relationship, goal, method, and overall (Miller & Duncan,
2004). A recent systematic review and meta-analysis of the PCOMS used in MFS identified 18 studies that showed an overall small, but significant effect size of the PCOMS, compared to control conditions \( (g = 0.27, \text{CI} [0.14, 0.41], p < .001) \) (Østergård, Randa, & Hougaard, 2020). The PCOMS was not significantly better at supporting not on track clients or preventing client deterioration. The 10 studies from psychiatric settings did not reveal a significant effect of the PCOMS \( (g = 0.10, \text{CI} [-0.03, 0.23], p = .144) \), however the eight studies from counselling settings had a significant effect \( (g = 0.45, \text{CI} [0.31, 0.59], p < .001) \).

It was argued that psychiatric clients are likely to have more severe mental illness compared to counselling clients, and psychiatric clients may have lower levels of motivation or difficulties in cognitive capacity to reflect on MFS. It is also more likely that psychiatric clients are resistant to standard psychotherapy and early progress is not to be expected.

There have been some questions and concerns about the validity and reliability of the ORS. One of the included studies in the review by Østergård et al. (2020) utilised a small sample size of 73 clients and explored the effects of therapy after three sessions (Seidel, Andrews, Owen, Miller, & Buccino, 2017). The results found effect sizes of 0.83 with the ORS, and 0.44 with the OQ-45. In this same study, a distressed community subsample of 28 clients achieved an effect size of 0.59 on the ORS between time periods one and three, without any treatment. Concerningly, the review by Østergård et al. (2020) showed seven studies comparing the ORS with another generic measure and found a difference in effect size of 0.11, with more positive effects shown on the ORS. The review questioned whether the ORS was overestimating levels of clinical change.

2.7.1 Measurement feedback systems with at risk populations. Varying outcomes of MFS are evident across different settings, including limited positive effects in psychiatric settings (Østergård et al., 2020). This indicates that MFS may have particular challenges with some client groups. For example, an RCT was conducted utilising MFS with clients
diagnosed with moderate to severe depression, with depressive symptoms measured by the Hamilton Depression Rating Scale (Guo et al., 2015). The two primary outcome criteria were a decrease of at least 50% in the measurement score and a score of seven or less. Results showed that the MFS group met the first criteria for 86.9% of clients, compared with 62.7% in the TAU group. Furthermore, there were higher rates of clients reaching a score of seven or less in the MFS group, with 73.8% compared to 28.8% of those in the TAU group. The MFS group showed higher rates of treatment adjustments, which were predominantly dosage changes and higher antidepressant dosages, compared to the TAU group. It was suggested that fine-tuning of medication to an optimal level using MFS was the primary reason for the improvements.

Another study focused on a highly distressed group, diagnosed with a range of mental illnesses in a public mental health system in Israel (Tzur Bitan et al., 2020). The study took place over six months, or earlier if a client was deemed to have completed therapy. The results showed no differences in treatment outcomes between the MFS group, compared to the TAU group. However, clients in the MFS group showed significant improvement in ratings of the therapeutic alliance with clinicians, while no significant change was seen in the TAU group. The sample consisted of highly symptomatic clients who had been treated in public psychiatric systems for approximately three years. Additionally, the sample’s baseline distress scores were three standard deviations above the mean of the non-clinical population and one standard deviation above the mean of the clinical population. While MFS supported changes in the therapeutic alliance, it is suggested that a longer time frame of treatment may be required using MFS, for enhanced clinical outcomes with this population group.

Interestingly, an RCT has been conducted exploring the use of MFS in a psychiatry emergency centre with clients presenting with acute and severe psychiatric problems (van Oenen et al., 2016). A naturalistic sample of 370 clients with a range of mental illnesses were
randomised into a TAU or MFS condition, with the MFS condition having feedback of session-by-session measurement scores provided to both clinicians and clients. Results showed that after 12 weeks there were no significant differences across the conditions. One explanation for these results is that the effect size in the TAU condition was relatively large, 0.71 after 12 weeks, and further improvement would have been difficult to achieve. Other explanations that were proposed were that feedback effects do not work as well with more severe psychiatric clients and that there is reduced capacity to reflect on feedback during times of crisis.

A systematic review and meta-analysis have been conducted exploring feedback informed approaches in psychological services, specifically with young people aged 10 to 19 years of age (Tam & Ronan, 2017). It was comprised of nine quantitative studies and three qualitative studies. Of the 12 studies, nine utilised the PCOMS and over 50% of studies were conducted in the USA. Results found the overall effect size was small, with Hedges’s g indexes of 0.20 for independent-groups trials, 0.32 for single-group trials corrected for time-effect bias, and 0.28 for all trials combined. It was concluded that the use of MFS with young people could provide beneficial outcomes similar to that shown in adult-based studies, while noting feedback effects to be small.

One of the first RCTs exploring the use of MFS with young people used the Symptoms and Functioning Severity Scale (Bickman et al., 2011). The measure was completed by the young person, the parent, and the clinician, to provide data from multiple perspectives. Results showed that clinicians who received weekly feedback of measurement scores had clients who improved faster than those without feedback, regardless of which version of the measure the clinician viewed. Feedback effects were shown to have modest effect sizes. Furthermore, there was a dose-response effect with the effect size increasing considerably when clinicians viewed more versions of the measure (young person, parent, clinician).
The use of MFS with young people is a developing research field and some of the latest studies have begun to expand the literature in this area. In a recent study comparing MFS effectiveness for adults compared with children and adolescents, the Y-OQ and OQ-45 were utilised (Dyason, Shanley, O'Donovan, & Low-Choy, 2020). Results found that MFS approaches were more effective than TAU benchmarks, with 50% of adults and 64% of children and adolescents significantly improved after psychotherapy using MFS. The study aimed to validate benchmarks for the OQ-45 and identify new benchmarks for the Y-OQ.

2.7.2 Clinician fidelity in measurement feedback systems. Clinician fidelity using MFS involves clinicians appropriately following all the evidence-based steps for MFS use in day-to-day practice (Lewis et al., 2019). The fidelity of clinicians has been proposed to play a significant role in feedback treatment effects and there are several different processes of MFS that are critical. Krägeloh et al. (2015) suggest the progressive categorisation of MFS processes: clinician-only feedback, feedback to clinicians and clients, encouragement of mutual measurement discussion, and availability of formalised mechanisms to guide such discussion. The most basic category of MFS use involves measurement information being provided to the clinician only, allowing treatment planning to be adjusted if indicated. The top-level of MFS use involves the full integration of all the processes, which includes measurement information provided to both the clinician and client, with a formal procedure for discussion and SDM that can inform treatment planning.

A systematic review targeting self-reported mental health measures in MFS aimed to explore which MFS processes were used across studies and what processes were related to treatment outcomes (Krägeloh et al., 2015). Twenty-five studies were included that described clinicians’ levels of use across the differing MFS processes. The review found only 11 studies used all processes of MFS, with nine finding positive effects and two finding partial positive effects. Studies that did not incorporate the fully integrated use of all MFS processes
had a significantly lower proportion of full or partial outcome effects. The results indicated that formalised clinician and client feedback, with structured guidelines for discussion and treatment planning, were most strongly associated with improved client outcomes.

A study found significant treatment effects using MFS with not on track clients when the clinician reported that they actually used the feedback, which included clinicians discussing feedback with clients, giving homework to clients, and using the feedback to determine when to terminate therapy (De Jong et al., 2012). Another study found that feedback to the clinician alone reduced the negative rates of change with not on track clients in short-term therapy, with a large effect size (0.91) (De Jong et al., 2014). This effect size increased to a very large effect size (1.28) in reducing negative rates of change when feedback was provided to both the clinician and not on track clients. In long-term therapy, however, only feedback to both clinician and client had a small positive effect size (0.29) on rates of change, and this was the same across both on track and not on track clients.

2.8 Implementation of Measurement Feedback Systems

2.8.1 Clinician characteristics and attitudes. The benefits of MFS are becoming more widely known, and there is now increased training in the use of MFS in clinical programs (Overington et al., 2015). However, clinicians typically have shown low rates of administration of routine outcome measures and there are multiple barriers to their use in clinical practice (Batty et al., 2013; Hatfield & Ogles, 2004; Johnston & Gowers, 2005). A survey completed by clinicians found a strong association between clinician attitudes and MFS use, with attitudes accounting for 16% of the variability in any MFS use, and 35% of the variability in frequent MFS use (Jensen-Doss et al., 2018). Specifically, general attitudes towards monitoring and feedback predicted whether clinicians ever used MFS. Some clinicians preferred to use their own clinical judgement instead of relying on outcome
measures, and some were not convinced of the effectiveness of MFS (Ross, Ionita, & Stirman, 2016). At the same time, some clinicians have reported a lack of knowledge around MFS and high levels of anxiety and discomfort using such measures with clients (Ionita, Fitzpatrick, Tomaro, Chen, & Overington, 2016).

Critically, clinicians have raised concerns around the clinical usefulness and practicality of measures used in MFS. They report that some measures can take too long to administer, lack timely feedback, or they find it challenging to access results (Ionita et al., 2016; Ross et al., 2016). The ability to administer measures electronically has reduced some of these concerns (Overington & Ionita, 2012). Other concerns are the relevancy of measurement items being fed back to the clinicians and clients, and a lack of measure sensitivity to change (Bickman, 2008; Happell, 2008; Kelley & Bickman, 2009). There is a strong association between clinicians’ attitudes towards measure usefulness and MFS use (Chung & Buchanan, 2019). Therefore, the choice of the measure used in MFS needs to be suitable for the specific target group and carefully considered (Becker-Haimes et al., 2020).

Jensen-Doss et al. (2018) explored clinicians’ attitudes to standardised measures and found that clinicians who were more positive about a measure’s clinical utility, treatment planning properties, and practicality were more frequently using routine outcome measures. The practicality of the measure was shown to be the strongest predictor of frequent measurement use. Conceptualised feedback intervention theory states that the value that a clinician puts onto MFS and the information provided is critical in its implementation success (Waldron, Loades, & Rogers, 2018). Therefore, it is suggested that strategies for shifting clinician attitudes towards outcome measurement and MFS is one of the first steps in setting up such systems.
Several clinician characteristics have been shown to be associated with MFS use. It has been found that clinicians who primarily use CBT, compared to clinicians using other interventions, are more likely to use outcome measures (Hatfield & Ogles, 2004). Furthermore, CBT clinicians are less likely to see MFS as potentially harmful and more likely to view outcome measures as clinically useful and practical, compared to clinicians of other therapeutic orientations (Jensen-Doss et al., 2018). It has been proposed that the structure of CBT is congruent with the monitoring and feedback used in MFS. Additionally, clinicians primarily working with adults were found to be more likely to be aware of and use MFS, compared to those working with other clientele groups, including young people (Ionita & Fitzpatrick, 2014). It was proposed that there was less availability of outcome measures and MFS for young people compared to adults.

Clinicians with higher-level degrees have been found to be more likely to use routine outcome measures, compared to those with lower-level degrees (Ionita & Fitzpatrick, 2014). It was theorised that clinical training programs of higher-level degrees might have incorporated the use of MFS in their coursework. Studies have found that clinicians in private practice were less likely to administer outcome measures, saw less benefit in monitoring and feedback, and were less likely to find measures clinically useful and practical, when compared to clinicians working in institutional settings (Hatfield & Ogles, 2004; Jensen-Doss et al., 2018). Furthermore, fewer years of professional experience has been shown to be one of the strongest predictors of outcome measure use (Jensen-Doss et al., 2018). Clinicians with fewer years of clinical experience are more likely to have more recently graduated from their trainee program, which may have incorporated MFS training and use.

Intriguingly, other clinician-related factors have also been shown to affect MFS treatment effects. Clinicians reporting higher levels of self-efficacy concerning using MFS had clients who progressed faster than clinicians with low self-efficacy (De Jong et al., 2012).
Clinicians who were more likely to trust feedback from sources external to their own opinion (low internal feedback propensity) had clients with faster rates of change compared to clinicians with a high internal feedback propensity. Clinicians with a strong focus on achieving success (promotion focussed) were more likely to achieve better outcomes using feedback, when compared to clinicians who focus on preventing failures (prevention focussed) (De Jong & De Goede, 2015).

2.8.2 Organisational supports. Service and organisational factors play a vital role in the successful implementation of routine outcome measures and MFS. When the use of progress monitoring is incorporated into workplace procedures and policies, clinicians tend to hold more positive views towards outcome measures and are more likely to administer them (Jensen-Doss et al., 2018). Importantly, organisational resources that incorporate technology to reduce administrative burden and increase the timeliness of feedback can help improve the use of MFS (Bickman, Douglas, et al., 2016). There can be concerns around too much paperwork and an extra burden on clients and clinicians when using outcome measures, while clinicians also report not having enough time to incorporate feedback measures into the session (Hatfield & Ogles, 2004). Additionally, there can be costs associated with using MFS, mainly those administered electronically, and this can deter some clinicians from using MFS when working privately or in small organisations (Overington & Ionita, 2012).

A trial of MFS in two youth clinics revealed that the clinic that provided more organisational supports engaged in MFS more often and had better treatment outcomes (Bickman, Douglas, et al., 2016; Gleacher et al., 2016). While both clinics had the same training protocols and barriers to using new technology systems, organisational supports targeting championing of MFS and day-to-day clinical supervision were seen to be the critical factors for successful implementation of MFS. The use of MFS has been shown to increase after initial training, however, a decline in use is usually seen later in practice.
(Persons, Koerner, Eidelman, Thomas, & Liu, 2016). Therefore, there are recommendations for ongoing clinical support, such as championing and supervision, to be provided to sustain MFS use (Connors et al., 2020). Champions are clinicians who actively associate themselves with evidence-based practice and promote positive attitudes and behaviours towards the implementation of a specific practice amongst peers (Lewis et al., 2019). While clinical support, such as supervision, can help provide objective guidance and expertise in MFS use (Gelkopf, Mazor, & Roe, 2020).

2.8.3 Processes and fidelity. MFS can be separated into several core processes: routine administration of outcome measures, clinician review of the data, clients’ review of the data, shared discussion of the data, and collaborative treatment planning informed by the data. However, not all studies report which type of process was implemented, and fidelity to the process is essential to gain the most substantial treatment outcomes (Lewis et al., 2019). MFS fidelity has rarely been evaluated and investigating the factors affecting the implementation of MFS across the various processes is recommended. The use of electronic systems can help monitor fidelity, by seeing when clinicians administer measures, access measurement graphs, and document client discussions of measurement data (Ross et al., 2016).

A systematic review on the PCOMS used in MFS identified 18 studies and results showed that none of the studies reported clinician adherence levels to the MFS model and competency levels using MFS (Østergård et al., 2020). The frequency the measures were used in sessions was only reported in four studies, seven studies reported on how many clients did not use the measures in all the sessions, and seven studies did not report on the in session use of the PCOMS. Nine studies reported the frequency of supervision, but usually described the time frames of supervision in broad categories of weekly, monthly, or regularly. Two studies investigated the effect of the PCOMS as a function of time of its
implementation. While, the first study showed no differences across treatment outcomes between early and late implementation phases of MFS (Davidsen et al., 2017), the second study found superior MFS treatment outcomes in later stages of MFS implementation, compared to earlier stages (Brattland et al., 2018). An explanation for this was that clinicians were able to become more effective with more use of MFS.

Ross et al. (2016) studied the implementation of a secure web-based platform MFS rolled out across Operational Stress Injury clinics in Canada. An electronic measurement system was provided with appropriate training and technical support. The study found 85% of clinicians administering measures, 78% reviewing results, and 65% discussing results with clients. There were no differences in barriers between users and non-users when administering measures, however, non-users perceived “burden” and “organisational supports” as more prominent barriers when reviewing measure results. “Burden” was also a more significant barrier for non-users when discussing measurement results with clients. These results suggest that there may be different barriers associated with the different processes of MFS.

2.8.4 Implementation with young people. The research around the implementation of MFS with young people has generally been congruent with adult literature. However, there are still a greater number of research papers and outcome measures targeted towards adult MFS (Ionita & Fitzpatrick, 2014; Tam & Ronan, 2017). There are some settings specific to young people in which MFS are increasingly being implemented. In a trial of MFS in a school setting, clinician perceptions of a measure’s psychometric qualities and assessment skills predicted a significant increase in both the administration of measures and feedback to clients, across the implementation period (Lyon et al., 2019). The study targeted clinician behaviour change, and clinicians who felt more positive about the reliability and validity of a measure were more likely to increase their MFS use over time. It was proposed that
increasing MFS knowledge with clinicians increases the responsiveness to MFS post-training and subsequent practice change.

Similarly, Waldron et al. (2018) explored the implementation of MFS in a United Kingdom CAMHS across six months and found that clinicians’ use of MFS showed small significant increases from the beginning to the end of the implementation period. Clinicians’ attitudes towards MFS did not significantly change over this period, however, it was suggested that MFS use increased due to clinicians’ MFS use being monitored and reviewed by supervisors. Thematic analysis of the qualitative results reported that clinicians were generally positive about MFS. Similar to the adult literature, there were concerns of funding bodies misusing the data and that it was management-directed rather than clinician-directed. Organisational supports that encourage clinicians’ self-directed use indicate more sustained MFS use over time and is preferred over management-directed MFS use (Sharples et al., 2017).

Clinician fidelity of MFS use has been shown to affect how young people view MFS. When the appropriate steps are followed in the use of MFS, young people’s perceptions of MFS are more positive, and this increases treatment outcomes (Mayworm, Kelly, Duong, & Lyon, 2020). Specifically, the way clinicians introduce, explain, and implement MFS impact how young people perceive it. Clinicians who do not utilise the useful parts of electronic MFS, such as the immediate feedback and graphic display of results, can limit the features of MFS that young people find appealing. This highlights the critical aspects of clinician fidelity in the implementation of MFS processes with young people.

2.9 Chapter Summary and Rationale for Current Research

In summary, there is a high prevalence of mental health issues for young people and, concerningly, high rates of deterioration when utilising mental health treatments (Warren et
al., 2009). Young people are also more likely to show higher treatment drop-out and missed appointments than adults, when accessing mental health supports (O’Brien et al., 2009). The use of MFS is now recommended to reduce clinician blindspots during treatment, support SDM and treatment planning, and is associated with improved treatment outcomes, particularly with those clients not on track (Carlier et al., 2012; Østergård et al., 2020; Shimokawa et al., 2010).

It is recommended that session-by-session mental health outcome measures are implemented into an electronic system to be used effectively in MFS (Hall et al., 2015). Specifically, diagnostically-generic outcome measures that are brief and clinically meaningful are urgently needed for young people aged to 12 to 25 years (Cleverley, Brennenstuhl, et al., 2020). Furthermore, there is now an understanding that statistically derived benchmarks for measures are a critical component of MFS, to allow clinicians to track clients’ progress against expected change (Dyason et al., 2020).

Multiple factors affect the implementation of MFS, including clinician characteristics and attitudes, and organisational supports. Furthermore, there is now emerging research exploring the implementation factors affecting the different processes of MFS use (Lyon et al., 2019). There is evidence to show that when all MFS processes are utilised, there are better treatment outcomes (Krägeloh et al., 2015). The next step in expanding this area of research should be to investigate the range of implementation factors and how they differentially impact the use of each MFS process (Lyon et al., 2019).

2.10 The Implementation of MyLifeTracker in headspace

MLT is a brief mental health outcome measure designed for young people aged 12 to 25 years. The measure was developed specifically for use in headspace through a literature review, consultation with clinical staff, clients, and the headspace youth reference group, and
pilot testing. MLT items measure the current self-reported quality of life in the following five domains of importance to young people: general wellbeing, day-to-day activities, relationships with friends, relationships with family, and general coping.

MLT was created to supplement other measures used in headspace, including the K10 and the clinician-rated Social and Occupational Functioning Scale (Rickwood, Mazzer, Telford, et al., 2015). MLT was designed for young people aged 12 to 25 years and, as discussed previously, there are no brief outcome measures specific for this age group. MLT was developed to be used every session, while also being meaningful to a range of young people, so the tool needed to be both brief and broad. It also needed to be implemented via an electronic format.

MLT is currently implemented via the electronic data system within headspace youth mental health centres throughout Australia. It is administered as part of a suite of other measures through a tablet device or computer before a young person attends each clinical session, sent via email, text message, or completed while they are waiting in reception. The young person’s self-reported results are instantly available to the clinician prior to the session in the form of a graph over time. The graph shows every session the young person has attended and can be viewed as a total MLT score or items can be individually viewed.

headspace is the Australian Government’s National Youth Mental Health Foundation, initiated in 2006 to provide early intervention in youth mental health for those aged 12 to 25 years (Rickwood et al., 2018). headspace centres offer holistic primary care services responding to mental health, alcohol and other drugs, physical health, and vocational concerns for young people. headspace centres have been implemented gradually across Australia, from 10 centres in 2007 to 110 centres in 2020.
The *headspace* data collection system routinely collects information from clinicians and clients. The system was initially implemented for quality improvement and program evaluation purposes (Rickwood et al., 2014), but is increasingly used as a clinical support tool. New clinicians to *headspace* are provided with an online orientation course, through the *headspace* learning program, that introduces the electronic data collection system and its uses at a national, centre, and client level. *headspace* clinicians are required to complete clinician-reported measures after each session, however, they are not required to use client-reported data in a mandated way, and they can choose whether and how they integrate the data into their clinical practice.

### 2.11 Research Aims

Based on the preceding arguments, the overarching purpose of the present body of work is to examine the development and implementation of youth mental health outcome measures for young people aged 12 to 25 years, to be used within MFS. This thesis employs a quantitative methodology to test and optimise MLT to be used within MFS, and investigate factors affecting its implementation across *headspace* services in Australia. There are four specific objectives with several aims and sub-aims for this research.

The first objective was to conduct a critical review of the existing literature on mental health outcome measures for young people aged 12 to 25 years. A systematic review of the literature was conducted, with only diagnostically-generic mental health outcome measures included. These consisted of measures of cognition and emotion, functioning, quality of life, and multidimensional factors. The main aim was to identify measures applicable to a general mental health population or used with a variety of specific mental health populations (rather than be unique to a particular mental disorder or condition). Specifically, young people were
targeted by including studies with a sample’s mean age between 12 to 25 years and excluding studies with samples that consisted of only participants aged 18 years and older.

To provide an understanding of how each measure had been used, data extraction focussed on several areas. A sub-aim of the review was to identify who used the measure and the following client demographics were extracted: age, gender, ethnicity, socioeconomic status, country, and diagnosis. Another sub-aim of the review was to explore how the measure was used in the study and the data extracted included: sample size, research design, setting, time of follow-up, reporter, measure change magnitude, and use in MFS. The systematic review identified no measures explicitly created for young people aged 12 to 25 years and only a few measures used in MFS.

The second objective was to test and validate a youth-specific mental health outcome measure. This research aimed to examine the psychometric properties of MLT used with young people aged 12 to 25 years, accessing support from headspace centres across Australia. It was hypothesised that the five items constituting MLT would represent a single factor when applying exploratory and confirmatory factor analysis. A multigroup confirmatory factor analysis was also performed to test measurement invariance across age and gender groups. Cronbach’s alpha was used to explore internal consistency, while Pearson’s correlation analysis was conducted to examine test-retest reliability. Concurrent validity was examined using Pearson’s correlation analysis, which explored the associations of MLT with standardised outcome measures of psychological distress, life satisfaction, wellbeing, and social and occupational functioning.

MLT was also tested for sensitivity to change and this quality in a measure is not commonly tested (Bentley et al., 2019). Sensitivity to change was examined using several methods, including effect sizes, RCI, and growth curve modelling. While effect sizes can
indicate the magnitude of change between the first and last sessions, RCI can provide a breakdown of proportions of young people who reliably improved, did not change, or deteriorated (Wise, 2004). RCI was also calculated for K10 scores to allow comparisons with MLT. Finally, growth curve modelling was estimated for MLT and this method can provide average rates of change and patterns of change (Singer & Willett, 2003). Validity, reliability, and sensitivity to change were further examined as a function of age group (12-14, 15-17, 18-21, 22-25) and gender (male/female).

The third objective was to optimise a youth-specific outcome measure to be useful in clinical practice. This research aimed to identify benchmarks for MLT within *headspace* centres that can be used in MFS. To be clinically meaningful, it is essential that measures have statistically derived benchmarks (Dyason et al., 2020). The first sub-aim was to determine MLT CSI that would show developmentally appropriate thresholds that would indicate if a young person presented in the clinical population, and if they moved into the non-clinical population during treatment (Jacobson & Truax, 1991). MLT CSI were determined using a clinical population, comprised of young people attending *headspace* centres, and a non-clinical population, comprised of a nationally representative community sample from across Australia. MLT CSI were also determined for each age group (12-14, 15-17, 18-21, 22-25) and gender (male/female). While being a distinct developmental group, there are still differences across the 12 to 25 year age range and by gender for mental health outcomes (Van Droogenbroeck et al., 2018).

Research has increasingly focussed on detecting clients who are at risk of deterioration using early warning systems that are derived from expected change trajectories (Crits-Christoph et al., 2015). Therefore, the second sub-aim was to determine expected change trajectories and early warning signals for MLT. Expected change trajectories were calculated for decile groups dependent on MLT baseline scores, as previous research has
shown that this accounts for the most variability in rates of change (Finch et al., 2001). Lastly, the third sub-aim was to provide a guide and clinical examples of how clinicians can use the statistically derived benchmarks for MLT in their clinical practice. Such frameworks are vital as they can increase positive clinician attitudes towards MFS and clinician fidelity to MFS use (Lewis et al., 2019).

The fourth objective was to explore the factors affecting the implementation of an outcome MFS in youth mental health settings. The main aim of this research was to investigate the factors that affect the implementation of MLT in MFS used across headspace centres. The first sub-aim of this study was to determine how much clinicians were currently using MLT across three MFS processes: look at MLT before session, use MLT in treatment planning, and provide feedback of MLT scores to clients. It is essential to break down these MFS processes, as there is emerging evidence that there are different predictors associated with clinicians’ use of different processes of MFS (Lyon et al., 2019). As part of the study, the second sub-aim was to examine clinician attitudes and perceptions towards MLT and MFS. This included examining clinician attitudes to MFS in general, attitudes to MLT specifically, and perceptions of organisational supports for MLT. This would help explore the acceptability and usefulness of MLT and clinician perceptions of headspace’s support for MLT use in MFS.

The third sub-aim of this study was to identify the strongest predictors of MLT use in each MFS process, specifically: clinician characteristics, attitudes to MFS in general, attitudes to MLT specifically, and perceptions of organisational support for MLT. Previous research has identified clinician characteristics increasing MFS use, including being a CBT clinician, fewer years of clinical experience, having higher-level degrees, and working in an institutional setting (Ionita & Fitzpatrick, 2014). Intuitively, clinicians who report increased benefits and reduced harm in MFS use are more likely to use MFS (Jensen-Doss et al., 2018).
Clinician attitudes towards a measure’s clinical utility, treatment planning properties, and practicality have been shown to increase MFS use. In addition to clinician-related factors, organisational supports were also included in the analyses, including resources, championing, clinical support, and training (Lewis et al., 2019). Examining the strongest predictors across the different processes of MFS use has practical implications for increasing clinicians’ fidelity to MFS use and the effectiveness of MFS.

The next four chapters of the thesis consist of four published articles that address these objectives and aims. Chapter III is a 2015 published article titled “A systematic review of mental health outcome measures for young people aged 12 to 25 years”. Chapter IV is a paper published in 2018 entitled “Development and validation of MyLifeTracker: A routine outcome measure for youth mental health”. This is followed by Chapter V that presents a 2020 published paper titled “A routine outcome measure for youth mental health: Clinically interpreting MyLifeTracker”. Chapter VI consists of an article also published in 2020 entitled “Factors affecting the implementation of an outcome measurement feedback system in youth mental health settings”.

Chapter III: A Systematic Review of Mental Health Outcome Measures for Young People Aged 12 to 25 Years

This chapter presents the first of four papers prepared during the PhD candidature. The paper is a systematic review aimed to identify the gap in existing mental health outcome measures used with young people and examine how these measures were being used clinically. Paper One has been peer reviewed and was published in *BMC Psychiatry* in November 2015. *BMC Psychiatry* is a Q1 journal in the Psychiatry and Mental Health category with an h-index of 88, a 2-year impact factor of 2.704, and a 5-year impact factor of 3.386. This paper has been cited in 45 other journal articles since publication. The format of the paper is in accordance with the publication guidelines for this journal.
Declaration of Co-Authored Publications

For use in theses which include co-authored publications. This declaration must be completed for each co-authored publication and to be placed at the start of the thesis chapter in which the publication appears, or as a preface to the thesis.

Declaration for Thesis Chapter III

DECLARATION BY CANDIDATE

In the case of Chapter III, the nature and extent of my contribution to the work was the following:

<table>
<thead>
<tr>
<th>Nature of Contribution</th>
<th>Extent of Contributions (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptualisation and design of the study, review of the literature, and write up of the article.</td>
<td>75%</td>
</tr>
</tbody>
</table>

The following co-authors contributed to the work:

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
<th>Contributor is also a UC student (Yes/No)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debra Rickwood</td>
<td>Conceptualisation and design of the study, review of the literature, and editing the article.</td>
<td>No</td>
</tr>
</tbody>
</table>

Candidate’s Signature  

18/10/2020  

Date

DECLARATION BY CO-AUTHORS

The undersigned hereby certify that:

1. the above declaration correctly reflects the nature and extent of the candidate’s contribution to this work, and the nature of the contribution of each of the co-authors.
2. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. there are no other authors of the publication according to these criteria;
5. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Location(s): University of Canberra

Signatures

Date  

19/10/2020
A systematic review of mental health outcome measures for young people aged 12 to 25 years

Benjamin Kwan1* and Debra J. Rickwood1,2

Abstract

Background: Mental health outcome measures are used to monitor the quality and effectiveness of mental health services. There is also a growing expectation for implementation of routine measurement and measures being used by clinicians as a feedback monitoring system to improve client outcomes. The recent focus in Australia and elsewhere targeting mental health services to young people aged 12–25 years has meant that outcome measures relevant to this age range are now needed. This is a shift from the traditional divide of child and adolescent services versus adult services with a transitioning age at 18 years. This systematic review is the first to examine mental health outcome measures that are appropriate for the 12 to 25 year age range.

Methods: MEDLINE and PsychINFO databases were systematically searched to identify studies using mental health outcome measures with young people aged 12 to 25 years. The search strategy complied with the relevant sections of the PRISMA statement.

Results: A total of 184 published articles were identified, covering 29 different outcome measures. The measures were organised into domains that consisted of eight measures of cognition and emotion, nine functioning measures, six quality of life measures, and six multidimensional mental health measures. No measures were designed specifically for young people aged 12 to 25 years and only two had been used by clinicians as a feedback monitoring system. Five measures had been used across the whole 12 to 25 year age range, in a range of mental health settings and were deemed most appropriate for this age group.

Conclusions: With changes to mental health service systems that increasingly focus on early intervention in adolescence and young adulthood, there is a need for outcome measures designed specifically for those aged 12 to 25 years. In particular, multidimensional measures that are clinically meaningful need to be developed to ensure quality and effectiveness in youth mental health. Additionally, outcome measures can be clinically useful when designed to be used within routine feedback monitoring systems.

Keywords: Youth, Young people, Mental health, Change, Routine outcome measure, Feedback

Background

An outcome measure in mental health care can be defined as a tool used to measure the effect on a person’s mental health as a result of health care intervention, plus any additional extra-therapeutic influences [1]. Specifically, outcome measures are quantitative indicators used at two or more points in time: baseline, post-intervention, discharge, or follow-ups [2, 3]. Routine outcome measurement, whereby the same outcome measure is used frequently at a number of time points, has been adopted in child and adolescent mental health services in Australia, New Zealand, Denmark, United Kingdom and Norway [4]. This push has been driven by an increasing emphasis on monitoring the quality and effectiveness of services [5, 6]. Routine outcome measurement reported at the service level enables decision making around funding of services, particularly at a government level where health
resources are limited and need to be distributed to achieve the best outcomes [4]. It is also essential as a component of ongoing service-level quality improvement. Importantly, routine outcome measurement improves clinical practice when it is part of a feedback monitoring system for clinicians [7]. When mental health measures are regularly provided to the clinician they can inform clinical decision making and enable the clinician to adjust treatment planning accordingly [8]. In adult mental health services feedback has been shown to increase accuracy of diagnosis, improve communication between client and clinician, enhance treatment monitoring, and help clients maintain positive effects for longer periods [9, 10]. For clients who are not improving or who are deteriorating during therapy, feedback systems can help improve outcomes [9]. Emerging research in youth mental health contexts suggests similar benefits of feedback monitoring systems for younger clients [11, 12].

To be useful, mental health outcome measures must be valid and reliable, sensitive to change, comparable across relevant client groups and service types, and meaningful to both clients and clinicians [6]. Fundamentally, outcome measures must be sensitive to change and be able to clearly convey the magnitude of change achieved [13]. However, measuring change is complex and needs to go beyond reporting statistical significance. Effect sizes and the timescale in which the change is evident are essential [14]. An increasingly used technique is calculating a measure of reliable change, which takes into account the reliability of the measurement instrument and has been proposed to provide a more accurate standard of meaningful change [15, 16]. Additionally, estimating clinical significance, which is distinct from statistical significance, has been recommended in mental health contexts. Change is clinically significant when a client moves from the dysfunctional to the functional range during therapy. This technique is not commonly used as it requires comparison populations and norms [17]. These metric are useful, however, as a client can be considered “recovered” when their outcomes show both reliable change and clinical significance [16]. Meaningful changes are also those that are of value and considered important by the client, family or clinician [18, 19]. Notably, quantifiable change can be different from perceived change, which means that it is important to determine outcome measures that are personally meaningful to clients [20].

Outcome measures need to be comparable over relevant client groups and treatment settings, and help inform initial case formulation and client prioritisation access. Outcome measures are increasingly designed to measure broad mental health status rather than assess symptoms associated with the diagnosis of specific mental disorders [21]. Specific measures may be required for diagnosis, but are not helpful when making comparisons between cases and services where differences in case mix exist [22]. Using specific diagnostic measures also means clinicians need to isolate a particular presenting problem at baseline to assess subsequent change. This presents challenges for the common situations when clients have comorbid mental health issues or their presenting issues change over the course of therapy [11, 23]. In contrast, measures of general mental health can be used in a range of mental health settings with different client characteristics, including public mental health agencies, private organisations, schools, and hospitals. Being generically relevant to a broad range of mental health presentations enables the measure to cater for clients with no disorder, such as those accessing prevention mental health programs, through to those with severe disorder, such as inpatient hospital clients [24]. It is important to note the role of outcome measures in epidemiological studies to track naturalistic change in non-clinical populations.

To be clinically useful, outcome measures need to be meaningful to clients and relevant to the areas in which they have treatment goals. Research with mental health service consumers shows that many measures are not particularly relevant to their situations and do not capture outcomes that are personally meaningful [25]. Determining an outcome measure that is applicable in both clinical work and service evaluation is challenging [26]. Mental health is a broad construct that comprises a number of different measurement domains [27]. These include measures that cover recovery, cognitive performance and emotional experience, ability to undertake daily activities and maintain interpersonal relationships consistent with development stage, and general life satisfaction and wellbeing [1, 28, 29]. Each domain has been recognised as providing a meaningful aspect of a client’s mental health status, but may vary in value for clinical use, service evaluation and epidemiological studies [27].

There is a long history of outcome measures for adult mental health services and for child and adolescent services, including both community-based and inpatient settings. In Australia, a comprehensive report on outcome measurement in community settings identified 136 measures, of which 31 were deemed most appropriate and being relevant for children and adolescents, adults or older persons [29]. The measures incorporate both client and clinician reporters, and parent reporter measures were available for children and adolescents [22]. Historically, outcome measures have either been targeted towards children and adolescents or adults, reflecting the traditional demarcations within the mental health care system [30]. For example, the Health of the Nation Outcome Scales (HoNOS) has two versions, one for adults aged 18 to 64 years and a child and adolescent
version (HoNOSCA) for those aged under 18 years [31]. Outcome measures specific to the youth transition period of adolescence and young adulthood are urgently needed due to recent changes in mental health service delivery specifically targeting this age range [32].

Reorienting mental health services to focus on young people is supported by understanding that they have the highest burden of mental illness across the lifespan, comprising 55% of the burden of illness for the 15 to 24 year old age group [33]. At least one in four young people aged 12 to 24 years experiences a mental health problem in any given year [34]. Research indicates that 75% of people suffering from a psychiatric disorder in adulthood experience onset by the age of 24 [35]. Of particular concern, however, young people are least likely to access support from mental health care organisations [32]. A systematic review of barriers and facilitators to mental health help-seeking in young people from qualitative studies identified the major barriers as problems recognising symptoms, a preference for self-reliance, and perceived stigma and embarrassment [36]. There is also a pervasive belief among young people that seeking help does not help [37]. Consequently, ways to ensure mental health support is effective, and perceived to be so, are essential to engage young people in services [38], and this requires being able to demonstrate meaningful outcomes from young people’s mental health service use [39].

Due to increased vulnerability to mental disorder during adolescence and early adulthood, the transition from child and adolescent to adult mental health services at the age of 18 years is extremely disruptive to effective mental health care; it undermines continuity of care at the time when this needs to be strongest [40]. Early intervention youth mental health initiatives are strongly promoted in Australia [41] and gaining momentum in many other countries [42]. Youth-focused service innovations focus on the importance of factors such as youth participation, shared decision making, and easy early access. This has led to the development of tools and supports aimed at engaging young people, such as age appropriate psychosocial and mental health assessments [43]. New methods of delivering mental health interventions to young people have also emerged, which include online and smart phone applications of counselling, self-help, assessment, and support groups [44].

Consequently, appropriate outcome measures are now required that are appropriate to young people’s developmental, social and emotional stages [45, 46]. The current study comprised a systematic review to identify appropriate mental health outcome measures for young people aged 12 to 25 years. Specifically, the review aimed to identify outcome measures that could be used for a broad range of mental health presentations and assessed mental health through global measures of cognition and emotion, functioning, quality of life and multidimensional factors (rather than focussed on specific diagnostic symptoms). The review aimed to explore how outcome measures have been used to track change, in what populations and settings they have been used, and whether they have been used as a feedback monitoring system to clinicians.

Methods

Search strategy

The search was conducted using the MEDLINE and PsychINFO databases, covering studies published since the inception of each database until the 9th June 2014. The search terms comprised four categories: young people, measures, mental health and change (see Table 1). These were combined using ‘and’ statements and searches were performed on article titles, abstracts and subjects. Additional studies were identified through hand searching the references of relevant studies and reviews. The search methodology and reported findings comply with the relevant sections of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [47]. See Additional file 1 for PRISMA checklist.

Eligibility criteria

The eligibility criteria included articles reporting global measures of mental health, used with a range of mental health populations for young people aged 12 to 25 years, and measuring change over time. Case studies, reviews, single study specific outcome measures and studies including participants with other medical conditions were excluded. To be included, studies had to:

- be written in English;
- include participants with a mean age in the range of 12 to 25 years;
- describe an outcome measure used as a general measure of mental health, including measures of

<table>
<thead>
<tr>
<th>Table 1 Search terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categories</td>
</tr>
<tr>
<td>Young person</td>
</tr>
<tr>
<td>Measures</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Mental health</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Change</td>
</tr>
</tbody>
</table>

*Is a wildcard character that may be used in place of any number of characters in a search word
emotion and cognition, functioning, quality of life and multidimensional mental health;
- report outcome measures tracking change over at least two measurement time points; and
- be applicable to a general mental health population or used with a variety of specific mental health populations (rather than be unique to a particular mental disorder or condition).

Additionally, the criteria excluded studies:
- of only adult or child participants;
- that were case studies or reviews;
- where participants had conditions related to physical health, developmental delays, neurological impairments, intellectual disabilities, learning disorders, situational stress/trauma and substance or alcohol dependence; and
- which had an outcome measure that was single study specific.

Data extraction
Following the database search, duplicates were firstly removed. Titles and abstracts were then screened and irrelevant studies removed. Full text articles of studies identified as possibly relevant for inclusion were then obtained and both authors inspected these against the eligibility criteria for inclusion. The database search was extensive, but authors of the published articles were not contacted to obtain further information to that published. Additional searching by name of each outcome measure identified in the review was not conducted as the aim was to identify outcome measures that met the eligibility criteria rather than identify every published article on the identified measures. Figure 1 shows the PRISMA flow diagram for study inclusion.

Relevant information from each article was entered into a spreadsheet that included: age, gender, ethnicity, socioeconomic status, country, diagnosis, sample size, research design, setting, time of follow-up, reporter, measure change magnitude, and use in feedback monitoring systems. The articles were then sorted into groups by the outcome measure(s) identified in the article. If more than one eligible measure was reported, the article was included under each relevant outcome measure group. Lastly, the outcome measures were categorised according to the major domains of cognition and emotion, functioning, quality of life or multidimensional.

Results
Search results
The search strategy identified 184 published articles covering 29 different outcome measures, with many articles...
identifying more than one measure. The key characteristics of each article by type of measure are summarised in Additional file 2. The outcomes comprised eight measures of cognition and emotion, nine of functioning, six that were quality of life, and six multidimensional mental health measures. The GAF, a measure of functioning, was the most commonly referenced measure overall. The most referenced measure of cognition and emotion was the CBCL; for quality of life, it was the SF-36; and the most referenced multidimensional measure was the HoNOSCA.

**Age range**

Figures 2, 3, 4 and 5 show the age range and mean age for each measure in each article. Of the 29 outcome measures, 22 were used in at least one study with a sample that ranged across the age 18 child/adult demarcation point. However, only 11 of these measures were used in samples that had mean ages in both the 12 to 17 and 18 to 25 year groups. These included the BPRS, GHQ-12, K10, SCL-90-R, YSR, CGAS, CGI-S, GAF, SOFAS, SF-36, and WHOQOL-BREF. The YSR and CGAS were used predominately in the under 18 year

---

**Fig. 2** Age range and mean of cognition and emotion measures
age range. It is important to note that none of the multi-dimensional measures were used in samples with mean ages both above and below 18 years. There were three measures used with young people below 18 years that had an adult countermeasure used at follow-up: the YSR (YSR), CGAS (GAS), and HoNOSCA (HoNOS).

Eight measures were used across the whole 12 to 25 year age range, comprising the BPRS, GHQ-12, K10, YSR, CGI-S, GAF, SF-36 and WHOQOL-BREF. The BPRS was predominately used with samples diagnosed with psychiatry and schizophrenia. The YSR was slightly modified in one article, with the term ‘kids’ changed to ‘young people’, so that it could be used across the broader 12 to 25 year age group, rather than just with those aged under 18 years. The GHQ-12 was only used in non-clinical samples, and mainly over longer time periods tracking naturalistic change. Consequently, there were only five measures deemed to be suitable for use across the whole 12 to 25 year age range and applicable to a variety of clinical and research settings and population groups: K10, CGI-S, GAF, SF-36 and WHOQOL-BREF.
Outcome measure reporter

Outcome measures can be reported by the client (self-report), parents/carers or clinician. The review identified 13 of the 29 measures as having more than one reporter. Eight measures were self-report by the young person, which were the GHQ-12, K10, YSR, YP CORE, and all quality of life measures except the KIDSCREEN-52 and WHOQOL-BREF. Eight measures were clinician reported, comprising the HoNOS and most measures of functioning. All articles that referenced the SOFAS did not note the reporter, but this measure is clinician assessed.

Outcome measures with options for all three reporters were the SDQ, SFSS, CGI-I, and Ohio Scales. Only one of the 13 articles for the CGI-I used all three reporters, and this study aimed to explore differences between reporter types. The CBCL was the only measure with a teacher report in one article. This measure also has parent and clinician reporter versions, and a self-report version, the YSR, in the

Fig. 4 Age range and mean of quality of life measures
same family of measures and used in a number of the same studies. The CBCL and YSR were kept distinct, however, due to the different age ranges they target.

**Population groups**
Out of the 29 outcomes measures, two were used specifically in non-clinical, community-based samples: the GHQ-12 and SWLS. The remaining 27 measures were used with various clinical participant samples, and 10 measures were used in both clinical and non-clinical samples: the CBCL, K10, SDQ, SCL-90-R, YSR, KIDSCREEN-52, SF-36, WHOQOL-BREF, BASC-2, and Y-OQ. It is important to note that all functioning measures and multidimensional measures, except the BASC-2 and Y-OQ, were used only in clinical samples.

**Intervention types**
All outcome measures were used in at least one trial or treatment interventions. Many also explored naturalistic change over time in the absence of an intervention, including the CBCL, GHQ-12, K10, SCL-90-R, YSR, KIDSCREEN-52, SF-36, SWLS, and WHOQOL-BREF.
The GHQ-12, KIDSCREEN-52, and SWLS were predominantly used to measure naturalistic change. No measures of functioning or multidimensional mental health were used to examine naturalistic change.

**Change magnitude**
The review determined whether the outcome measures were used to assess change using tests of significance, effect size, reliable change and clinical significance. All but one outcome measure (SWLS) reported change magnitude over time. There were 28 measures that reported tests of significance and 17 included effects sizes. Outcome measures showing small to medium effect sizes included the BPRS, CBCL, K10, SCL-90-R, SFSS, YSR, KIDSCREEN-52, SF-36 and YP CORE. Measures showing medium to large effect sizes included the SDQ, CAFAS, CGAS, CGI-S, GAF, EQ-5D, YQOL-R and HoNOSCA. Effect sizes of small, medium and large were based on Cohen’s d of 0.2, 0.5, and 0.8, respectively.

Of the five measures identified suitable for use with the whole 12 to 25 year range, the K10, CGI-S, GAF and SF-36 reported effect sizes. The K10 was used in one study involving 36 non-clinical participants comparing two online coping programs and a control over nine weeks. There was a significant main effect over time for all three groups, with a small effect size [48]. The CGI-S was used with 20 participants for treatment of anxiety, showing a large effect size over 14 weeks of treatment [49]. A study used the GAF with 74 psychiatric outpatient participants with a range of disorders being treated with a Mindfulness-Based Stress Reduction (MBSR) program compared to Treatment as Usual (TAU). After post treatment (8 weeks) and follow-up (3 months) a large effect size was evident for the MBSR group compared to a small decline in the TAU group [50]. Lastly, the SF-36 was used with 63 participants being treated for first-episode mania, demonstrating small effect sizes on both mental and physical component scores after 6, 12 and 18 months [51].

Only seven of the measures were analysed using a reliable change index, which included the SDQ, SCL-90-R, YSR, CGAS, GAF, Ohio scales, and Y-OQ. The SDQ, SCL-90-R, and CGAS were used in randomised trials which showed reliable change index cut-off comparisons between intervention and control groups. The SDQ, used in a randomised trial of Acceptance and Commitment Therapy (ACT) compared to TAU, showed a reliable improvement for 26 % compared to 0 % at post treatment, respectively [52]. In a randomised trial reporting reliable change using the SCL-90-R, MBSR showed 59 % of participants with no change and 41 % improved, while TAU showed 10 % worsened, 62 % had no change and 27 % improved [50]. The CGAS was used in a study of female Apache American Indians with depression, to measure outcomes for a cognitive-behaviour based program versus an education support program. Differences in reliable change between the two interventions were reported at post intervention (8 weeks), 12 weeks, 20 weeks and 32 weeks [53].

Five outcome measures used tests of both clinical significance and reliable change; namely, the SCL-90-R, YSR, GAF, Ohio Scales, and Y-OQ. The GAF was used in a young adults’ counselling centre with 78 Swedes aged 16 to 23 years with a range of mental health disorders. A reliable improvement was calculated to be an increase of at least 10 points on the GAF. At post treatment, with mean length of treatment being 11 months, 52 % of participants showed reliable improvement while 48 % showed no change. Additionally, 31 % demonstrated clinically significant improvement [54]. A study using the Y-OQ in a school-based mental health treatment program reported both reliable change and clinical significance to conclude that 45 % of clients had “recovered” by meeting both criteria [55].

**Follow-up time frame**
The systematic review extracted follow-up time frames for the outcome measure studies, and categorised these as: short-term (0–6 months), medium-term (over 6 months-1 year), and long-term (over 1 year). Many measures were used across all three time frames. Measures used only in a short-term time frame included the SFSS, BACC-2, and Ohio Scales. The CGI-I, CGI-S, YQOL-R, and Y-OQ were primarily used in short-term time frames but did have some variation. Two measures primarily used in a long-term time frame were the KIDSCREEN-52 and GHQ-12. A small number of studies reported routine use of outcome measures, whereby the same measure was used at multiple time points: SDQ, SFSS, and Y-OQ.

**Sample demographics**
The outcome measures were all used with a range of sample demographics, according to gender, ethnicity and socioeconomic status. All 29 outcome measures were applied in equivalent ways for males and females; only one study using the K10 identified a participant that was transgender.

The majority of studies did not report ethnicity and, for those that did, there was little consistency. There were five measures that only reported primarily Caucasian samples, but no further information on what this broad category comprised, which were the VFI, SFSS, EQ-5D, BACC-2, and Ohio Scales.

In most studies, the socioeconomic status of the sample was not reported. Outcome measures that reported being used in a lower socioeconomic sample included
the CRCL, GHQ-12, SDQ, YSR, CGAS, CIS, GAF, SOFAS, VFI, and SWLS.

Feedback systems
The review identified three outcome measures used routinely, however only two of these measures were used as part of a feedback monitoring system, the SDQ and SFSS. In each case, these measures were used repeatedly to provide routine feedback to the clinician on the young person's outcomes. No functioning, quality of life or multidimensional measures were used as a feedback monitoring system.

The SDQ was specifically adapted in one study to be able to be used routinely in a feedback system. This session by session measure (SxS) was used to examine treatment effects using the feedback monitoring system during TAU over a year. Participants were recruited from Child and Adolescent Mental Health Services outpatient clinics, aged 11 to 19 years with a range of mental health disorders. The young clients and parents reported the SxS measure, which was fed back to clinicians and discussed with the young client. Results showed statistically significant changes after a year on the CGAS and young person reported SxS, however, no statistically significant changes in the HoNOSCA and parent reported SxS [56].

The SFSS was used in a randomised cluster controlled trial comparing weekly feedback versus no feedback monitoring system with young people being treated for a range of mental health disorders. Participants were recruited through a private health organisation, were aged 11 to 18 years and participated in the study for a mean time of 16.5 weeks. The SFSS used young person, parent and clinician reporters. Client participants with clinicians who received feedback on the SFSS improved faster than those with no feedback. Feedback effect sizes were small, being 0.18, 0.24, and 0.27 for reports from young people, clinicians, and parents, respectively. There were stronger effects when clinicians viewed multiple reporter sources; that is, from young persons, parents and clinician [12].

Discussion
This systematic review identified 29 mental health outcome measures, reported in 184 articles examining change in mental health status for young people aged from 12 to 25 years. It is the first review to examine outcome measures specifically across this age range. Prior reviews have reflected the traditional mental health service system demarcation by focussing on outcome measures used with either children and adolescents or with adults [22, 29]. The focus of the review was on general mental health outcome measures, rather than disorder-specific symptom measures, consistent with recent research highlighting the need to measure outcomes across comorbid conditions, changing presenting problems, and different client types and settings [22, 29].

Age range appropriate measures
There were eight outcome measures identified as being used across the whole 12 to 25 year age range. These were the BPRS, GHQ-12, K10, YSR, CGI-S, GAF, SF-36 and WHOQOL-BREF, but none are developed specifically for this target age range. Three measures are considered to be less appropriate for general youth mental health, namely: the BPRS, because it is used primarily with psychosis; the YSR, as it was specifically designed for clients under the age of 18 years; and the GHQ-12, which is mainly used with non-clinical samples to track naturalistic change. This leaves five measures deemed suitable for use across the whole 12 to 25 year age range and applicable to a variety of clinical and research settings and population groups: K10, CGI-S, GAF, SF-36 and WHOQOL-BREF. A discussion of their strengths and weaknesses follows.

The K10 was developed by Kessler and colleagues as a measure of non-specific psychological distress. It is a 10-item self-report measure which asks clients about symptoms of anxiety and depression in the past four weeks. The K10 has been widely used as a measure of mental health status in population surveys as well as an outcome measure in primary care settings. It demonstrates strong validity, excellent reliability and has been shown to be sensitive to change [29]. No studies have examined the feasibility of the K10 as a routine outcome measure; however, it has been noted as easy to use, brief and is one of the key outcome measures for the Better Outcomes in Mental Health Care Initiative in Australia [29]. In the current review, the K10 was shown to be used with clinical and non-clinical samples, tracking both treatment effects and naturalistic change. Change in the K10 was reported mainly with tests of statistical significance and a small effect size was demonstrated in one study.

The CGI-S is a brief clinician-rated global measure of current severity of the client's symptoms and functioning. The CGI-S is one-item asking the clinician, in their clinical experience, how mentally ill the client has been over the past week from "normal" to "extremely ill" [57]. The CGI-S has been shown to be sensitive to change, showing similar change to the HoNOS. It has been identified as suitable for routine use due to its brevity and ease of administration [58]. However, there are questions about its validity and reliability and efforts have been made to improve its psychometric properties [59]. In the current review, the CGI-S was used only with clinical samples, and change was reported using statistical significance and effect size, revealing large statistical effects.
The GAF is a clinician-rated scale giving a measure of overall psychiatric disturbance integrating three dimensions of functioning: psychological, social and occupational. It is a single-item measure on a 100-point scale divided into 10-point intervals [60]. It has shown good construct and concurrent validity, but questions have been raised over its content validity. Inter-rater reliability can be low, particularly in routine clinical use [29]. It is sensitive to change when correlated with change in the Positive and Negative Syndrome Scale (PANSS) [61]. The GAF is brief, easy to use and reliability can be increased with minimal training, which makes it more acceptable in routine clinical settings [29]. In the current review, the GAF was the most frequently referenced measure, was used in only clinical samples, and showed large effect sizes and both reliable and clinically significant change. The GAF was included in the revised third and fourth editions of the Diagnostic and Statistical Manual (DSM), but removed from Version 5 in favour of the World Health Organization Disability Assessment Schedule 2.0. The DSM-5 Task Force decided that the GAF was not an adequate assessment of psychiatric functional impairment due to its lack of conceptual clarity, the need for separate assessment of severity and disability, questionable psychometrics in routine practice, and the need for specific training for proper routine clinical use [62].

The SF-36 is a multipurpose, self-report, short-form health survey containing 36 items grouped under eight scales: physical functioning, role limitation due to physical functioning, bodily pain, general health, vitality, social functioning, role limitation due to emotional problems and mental health. The eight scales can be summed into physical and mental health summary scores. The SF-36 has been used with a range of mental disorders and physical diseases, and a variety of treatments. It has been shown to be valid, reliable, sensitive to change, brief and easy to use [29]. In the current review, it was the most referenced measure of quality of life, when including its shorter 12-item version. The SF-36 was used in both clinical and non-clinical settings, over short to long-term time frames, and showed small to medium effect sizes.

The Australian WHOQOL-BREF comprises 26 items measuring broad domains of physical health, psychological health, social relationships and environment over the last two weeks. It has good validity, reliability and sensitivity to change, however, has been suggested to be more appropriate for use at a population level [29]. In the current review, it was used primarily with clinical samples over short time frames, although one study used a larger non-clinical sample tracking naturalistic change. The WHOQOL-BREF has both young people and clinician reporters, however, self-report is recommended if the client has sufficient ability to complete the measure.

These five outcome measures were used effectively in studies of samples spanning the 12 to 25 year age range, even though they were originally developed for use with adults. None of these measures has been tested specifically for its clinical utility or psychometric properties for the youth age range. The current review did not identify any outcome measures developed specifically for the adolescent and young adult demographic. While these five measures seem promising, further tests of psychometrics and clinical utility are needed.

Despite the lack of targeted measures, there were 22 out of 29 outcome measures identified in the review that were used in at least one study with a sample that ranged across the 18 years of age mental health service system demarcation point. These included measures that were originally developed to be used with young people only up to the age of 18 years, such as the CBCL, YSR, and KIDSCREEN-52 [63–65]. This reveals the need for specifically developed and targeted measures for young people. There are major developmental changes that occur for young people from the ages of 12 years, around the time of the onset of puberty, to 25 years, which is well into adulthood [66]. It is highly likely that useful measures for this age range would need some clearly defined flexibility to accommodate developmental changes, particularly in areas of psychosocial functioning such as intimate relationships, education and work.

Type of reporter
Outcome measures can be self-report, clinician report or reported by relevant others (such as parents or teachers), and these different perspectives are all important for treatment [46]. In particular, self-report measures are essential for youth, to recognise their growing maturity and independence and engage them in their own treatment progress.

The place of parent reports may need further consideration, however, the current review identified very little use of parent reporters across the 12–25 age range, and only for children and adolescents [22]. New models of youth-focused care recognise the critical role of family, and parent reports may be relevant for clients up to 25 years of age [42], by providing another source of insight, particularly around changes in behavioural difficulties [67]. Careful attention would, however, need to be given to consent and confidentiality issues [66, 68].

Tracking change
All the outcome measures identified in this systematic review were used to track change over time. There were eight measures used primarily within a six month period, suggesting they might be more sensitive to
change in a relatively short time frame. In contrast, the KIDSCREEN-52 and GHQ-12 were used predominately in longitudinal population studies. Only three out of 29 outcome measures reported being used routinely at multiple time points: SDQ, SFSS, and Y-OQ. Routine use of outcome measures is a necessity when used as a feedback monitoring system, and this was demonstrated in studies using the SDQ and SFSS.

Only seven outcome measures were used to report reliable change, and only five of these also reported clinically significant change. This is concerning as studies have shown that reliable and clinically significant are more clinically meaningful change measures for mental health research [13]. These methods were designed to account for measurement error and clinical thresholds, requiring change to be statistically reliable and demonstrate movement from a dysfunctional to the functional population distribution [15]. Using these criteria, individuals can be classified into the outcome categories of recovered, improved, unchanged, or deteriorated, which are meaningful and interpretable categories [16]. However, it should be noted, that calculations of reliable change and clinical significance produce more conservative change results than other approaches [16, 69]. Further, in an early intervention context, clinical significance may not be appropriate as most clients may not present in the dysfunctional range to start with. In these contexts, clinical deterioration should be monitored, however, to determine whether clients change from the functional to the dysfunctional distribution, indicating need for higher levels of intervention. More research is needed in this area to determine optimal change indices for youth outcome measures.

Routine feedback
There has been an increase in demand for outcome measures to be used as a feedback monitoring system for clinicians [8]. Very few outcome measures were identified in the current systematic review that were used in this way, and these were designed for children and adolescents under 18 years [70]. The SDQ used young person and parent reporters and this information was fed back to clinicians to discuss with the young person. Treatment as usual with SDQ feedback showed statistically significant change on the CGAS post treatment, however, the study did not have a comparison group so it was unknown whether the change was due to the feedback, treatment as usual or the combination of both [56]. The SFSS study used young person, parent and clinician reporters and this information was fed back to clinician, but the study did not specify if this information was fed back to the young person. Feedback was found to improve client change, and this was heightened when feedback came from multiple sources [12]. Multiple feedback sources can provide different change perspectives of value to the clinician and young person client [46].

Of special note, the study using the SDQ within a feedback monitoring system showed that the measure had to be modified to be used in this way [56]. This suggests the possibility of other measures being modified or adapted to be used routinely. There are, however, several barriers to routine feedback, which may account for the small number of measures identified here [2, 71]. These include constraints around time, resources and training needed, and perceived lack of clinical utility [72]. There are likely to be additional barriers for young people as clients, as they are a unique client group with higher dropout rates, are often referred by parents or teachers rather than being self-referred, and have different goals for therapy and therapeutic expectations compared with adults [20].

Limitations
A thorough search strategy was employed in this systematic review and it identified a large number of outcome measures and studies, but it is possible that relevant measures were missed. Notably, article authors were not contacted for additional information and the methodology excluded articles that were not written in English, meaning measures used specifically in other cultures were excluded. The eligibility criteria also excluded articles pertaining to participants with other health conditions, including substance use and situational stressors. This was done partly to make the review more manageable, but may have excluded relevant measures. Only two databases were used in the search strategy, MEDLINE and PsychINFO, although these are the most commonly used in systematic reviews of mental health outcome measures [22, 45, 73]. Together, the database yielded an initial 11920 articles, which was filtered to a comprehensive 184 studies, identifying 29 outcome measures. Nevertheless, some measures, especially those not often used for research purposes and primarily used in clinical practice, may have been missed.

In particular, some popular outcome measures were not identified via the final criteria, including the Depression Anxiety Stress Scale (DASS) and the Outcome Rating Scale (ORS). The DASS is a self-report measure which comes in a 21 or 42-item version [74]. It is commonly used as individual scores for depression, anxiety and stress and, therefore, was excluded as measuring specific mental health conditions. The ORS is an outcome measure developed as a brief alternative to the Outcome Questionnaire 45.2 (OQ-45.2) [75]. The Y-OQ, which was included in this review, also comes from the same family of measures. There is a growing body of research around the ORS, particularly regarding its use as a feedback
monitoring system for clinicians [76]. However, in this review, it was excluded as it was unique to only one study with young people aged 12 to 25 years [77].

Conclusions
Mental health outcome measures are essential for quality assurance and monitoring the effectiveness of services, and for tracking longitudinal health trends across time [5, 6]. Although this review identified a large number of measures used with young people aged 12 to 25 years, only eight were used across this whole age range, each with strengths and weaknesses. Overall, the review found no measures designed specifically for young people. There is a growing push for outcome measures to be routinely used as feedback monitoring systems, and to determine clinically meaningful change [7, 20]. Only two measures were identified here as being used in this way and this is an area of particular research need for youth mental health because of the potential for such an approach to benefit clients [12]. Future research should focus on development of mental health outcome measures designed specifically for young people aged 12 to 25 years to accompany changes in mental health services that target this age range. The measures should be sensitive to reliable and possibly clinically significant change that is meaningful to young people, and also suitable for routine use as feedback to clinicians and young people themselves. This will provide services with age-appropriate measures with better clinical utility and comparative usefulness to drive delivery of the better mental health outcomes for young people, who have such a heightened need for early and effective mental health care.

Additional files

Additional file 1: PRISMA checklist. (PDF 192 kb)

Additional file 2: Mental health outcome measures used with young people 12 to 25 years [78–248]. (XLSX 46 kb)

Abbreviations
ACT: Acceptance and Commitment Therapy; ASR: Adult Self-Report; BASC-2: Behavioural Assessment System for Children-2; BPAS: Brief Psychiatric Rating Scale; CAFAS: Child and Adolescent Functional Assessment Scale; CBCL: Child Behaviour Check List; CGAS: Children’s Global Assessment Scale; CGI-I: Clinical Global Impressions-Scales-Improvement scales; CGI-S: Clinical Global Impressions Scales-Severity of Illness; CIS: Columbia Impairment Scale; DASS: Depression Anxiety Stress Scale; DSM: Diagnostic and Statistical Manual; EQ-5D: EuroQol; GAS: Global Assessment Scale; GHQ-12: General Health Questionnaire; GAF: Global Assessment of Functioning; GAS: Global Assessment Scale; GRQ: General Health Questionnaire; HCNSCA: Health of the Nation Outcome Scale; HoNOSCA: Health of the Nation Outcome Scale for Children and Adolescents; K10: Kessler Psychological Distress Scale; MBSR: Mindfulness-Based Stress Reduction; QoC-45.2: Outcome Questionnaire 45.2; ORS: Outcome Rating Scale; PANSS: Positive and Negative Syndrome Scale; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; SCL-90-R: Symptom Checklist 90 Revised; SDQ: Strengths and Difficulties Questionnaire; SF-12: Medical Outcomes Study (MOS) 12-item Short Form Health Survey; SF-36: Medical Outcomes Study (MOS) 36-item Short Form Health Survey; SFOS: Symptom and Functioning Severity Scale; SOFAS: Social and Occupational Functioning Assessment Scale; SWLS: Satisfaction with Life Scale; SxS: Session by session measure; TAU: Treatment as Usual; VFI,: Vanderbilt Functioning Index; WHOOQ-45.2: World Health Organisation Quality of Life Instrument-Brief; Y-OQ: Youth Outcome Questionnaire; YASR: Young Adult Self-Report; Y-0Q-30: Youth Outcome Questionnaire-30; YP CORE: Young Persons Clinical Outcomes for Routine Evaluation questionnaire; YQOL-R: Youth Quality of Life Instrument-Research Version; YSR: Youth Self-Report.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
BK designed and undertook the systematic review and drafted the article. DJR supervised the design, reviewed the results, and revised the article. All authors read and approved the final version of the manuscript.

Authors’ information
Benjamin Kwan is undertaking a PhD in Clinical Psychology in the Faculty of Health at the University of Canberra. Dr Debra J Rickwood is Professor of Psychology in the Faculty of Health at the University of Canberra and Chief Scientific Advisor to headspace The National Youth Mental Health Foundation.

Acknowledgements
None.

Author details
1Faculty of Health, University of Canberra, Kiirinari Street, Bruce, ACT 2601, Australia. 2Headspace National Youth Mental Health Foundation National Office, 485 La Trobe Street, Melbourne, VIC 3000, Australia.

Received: 20 April 2015 Accepted: 27 October 2015
Published online: 14 November 2015

References


doi:10.1097/CHI.0b013e3181a0eac.

doi:10.1136/adc.2010.188755.


doi:10.1097/CHI.0b013e3182474239.


Chapter IV: Development and Validation of MyLifeTracker: A Routine Outcome Measure for Youth Mental Health

Chapter IV presents the second paper from the research and aimed to examine the psychometric properties of MLT, a routine outcome measure for youth mental health settings that was implemented in *headspace* youth mental health centres throughout Australia. Paper Two has been peer reviewed and was published in *Psychology Research and Behavior Management* in April 2018. *Psychology Research and Behavior Management* is a Q2 journal in the Psychiatry and Mental Health category with an h-index of 25 and an impact factor of 2.030. This paper has been cited in six other journal articles since publication. The format of the paper is in accordance with the publication guidelines for this journal.
Declaration of Co-Authored Publications

For use in theses which include co-authored publications. This declaration must be completed for each co-authored publication and to be placed at the start of the thesis chapter in which the publication appears, or as a preface to the thesis.

Declaration for Thesis Chapter IV

<table>
<thead>
<tr>
<th>Nature of Contribution</th>
<th>Extent of Contributions (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptualisation and design of the study, review of the literature, data analysis, and write up of the article.</td>
<td>75%</td>
</tr>
</tbody>
</table>

The following co-authors contributed to the work:

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
<th>Contributor is also a UC student (Yes/No)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debra Rickwood</td>
<td>Conceptualisation and design of the study, review of the results, and editing the article.</td>
<td>No</td>
</tr>
<tr>
<td>Nic Telford</td>
<td>Conceptualisation and design of the study, review of the results, and editing the article.</td>
<td>No</td>
</tr>
</tbody>
</table>

Candidate’s Signature

Date: 18/10/2020

DECLARATION BY CO-AUTHORS

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate’s contribution to this work, and the nature of the contribution of each of the co-authors.
(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
(4) there are no other authors of the publication according to these criteria;
(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Location(s): University of Canberra

Signatures

Date

19/10/2020

Date

20/10/2020
Development and validation of MyLifeTracker: a routine outcome measure for youth mental health

Benjamin Kwan¹
Debra J Rickwood¹,²
Nic R Telford²

¹Faculty of Health, University of Canberra, Bruce, ACT, ²headspace National Youth Mental Health Foundation, Melbourne, VIC, Australia

Purpose: Routine outcome measures are now being designed for session-by-session use, with emphasis on clinically meaningful items and sensitivity to change. Despite an increasing mental health service focus for young people aged 12–25 years, there is a lack of outcome measures that are designed to be used across this age group. Consequently, MyLifeTracker (MLT) was developed as a brief mental health outcome measure designed for young people for routine use. It consists of the following five items targeting areas of importance to young people: general well-being, day-to-day activities, relationships with friends, relationships with family, and general coping.

Participants and methods: The measure was tested with 75,893 young people aged 12–25 years attending headspace centers across Australia for mental health-related issues.

Results: MLT showed a robust unidimensional factor structure and appropriate reliability. It exhibited good concurrent validity against well-validated measures of psychological distress, well-being, functioning, and life satisfaction. The measure was further demonstrated to be sensitive to change.

Conclusion: MLT provides a psychometrically sound mental health outcome measure for young people. The measure taps into items that are meaningful to young people and provides an additional clinical support tool for clinicians and clients during therapy. The measure is brief and easy to use and has been incorporated into an electronic system that routinely tracks session-by-session change and produces time-series charts for the ease of use and interpretation.

Keywords: MyLifeTracker, youth mental health, routine outcome measure, routine outcome monitoring, adolescent and young adult

Introduction

Routine outcome measurement (ROM) is now widely used in mental health services across Australia, New Zealand, Netherlands, Canada, Germany, the USA, Norway, and the UK.¹ This is in response to calls for greater quality assurance and demonstration of service effectiveness. Increasingly, decisions regarding government funding are based on evidence of cost-effective mental health outcomes.² Furthermore, there is expanding implementation of session-by-session ROM,³ whereby clients complete outcome measures at each session, targeting areas of cognition and emotion, functioning, and quality of life, with the aim of quantitatively informing clinicians about their clients’ progress.⁴

Some concerns are consistently raised by stakeholders about the use and implementation of ROM. Studies have shown typically low completion rates of ROM, which
suggestions a number of barriers in their clinical application.\textsuperscript{5–7} Clinicians have identified practical barriers, such as constraints around time and lack of training provided for the use of ROM.\textsuperscript{8–10} There are concerns around the lack of timely feedback of outcomes to clinicians, measures being meaningless to clients, and measures lacking sensitivity to change.\textsuperscript{11,12} Additionally, lack of administrative support and appropriate information technology support reduces the feasibility and acceptability of ROM.\textsuperscript{13} Consequently, services find it challenging to identify appropriate outcome measures for both clinical work and service evaluation.\textsuperscript{14}

Young people, clinicians, and parents have identified through qualitative research on session-by-session ROM that measures need to be brief, sensitive to change, and meaningful to both young people and clinicians.\textsuperscript{10,15} Notably, measures administered electronically using computer devices were found to be engaging to young people, and timely feedback using computer-enabled visual representations to display outcomes was useful to both young people and clinicians; electronically administered outcome measures are now being designed for routine use, with increased focus on measure brevity and sensitivity to change.\textsuperscript{16} Training and support tools for clinicians are also being developed for the use of ROM.\textsuperscript{16}

A number of measures have been designed for session-by-session use including the Outcome Rating Scale,\textsuperscript{17} Outcome Questionnaire-45 (OQ-45),\textsuperscript{18} Youth Outcome Questionnaire (Y-OQ),\textsuperscript{19} Clinical Outcomes in Routine Evaluation-Outcome Measure,\textsuperscript{20} Session-by-Session Strengths and Difficulties Questionnaire (S × S-SDQ),\textsuperscript{21} and Symptoms and Functioning Severity Scale (SFSS).\textsuperscript{22}

Session-by-session ROM provides clinicians with regular feedback on their clients’ progress, which enables appropriate decision making during treatment and can be used as a therapeutic tool to discuss progress in-session with a client.\textsuperscript{11} A growing evidence base shows that ROM can improve communication between clinician and client, reduce client dropout, increase the accuracy of diagnosis, maintain positive outcomes for longer periods, and improve outcomes for those who are not improving or who are deteriorating.\textsuperscript{4,23–25} This suggests that ROM feedback could be particularly beneficial for population groups who are difficult to engage in mental health services.\textsuperscript{26}

A developmentally distinct group who have the highest rates of mental illness but the lowest levels of help-seeking are adolescents and young adults.\textsuperscript{27–29} Collectively termed “young people”, there is an increasing shift away from the historical categorization of adulthood starting at the age of 18 years, as a number of complex changes occur socially, physically, neurologically, and psychologically up to the age of 25 years.\textsuperscript{30} Traditional mental health services are organized around a demarcation between “Child and Adolescent Mental Health Services” (CAMHS) and a major transition to “Adult Mental Health Services” at the age of 18 years. It is now well understood that this service dichotomy is particularly inappropriate for mental health services, as it undermines continuity of care at a time of greatest need and heightened vulnerability.\textsuperscript{31}

Mental health reform in Australia has reoriented mental health services to focus on early intervention for young people aged 12–25 years, and this approach has been gaining momentum in other countries.\textsuperscript{31,32} headspace is Australia’s Youth Mental Health Foundation, which was initiated in 2006 to provide evidence-based early intervention mental health services to young people, implementing a model that breaks down the barriers to help-seeking.\textsuperscript{33} Similarly, Jigsaw was a setup to strengthen the community’s capacity to support young people aged 12–25 years in Ireland and there are related initiatives burgeoning worldwide.\textsuperscript{34}

With the current strong focus on service innovation in youth mental health, researchers are exploring how to measure and identify therapeutic change for young people.\textsuperscript{35–37} A systematic review of mental health outcome measures for young people aged 12–25 years did not identify any measures designed specifically for this age range.\textsuperscript{37} The review focused on general mental health outcome measures, rather than single mental disorder-specific measures, and identified the target age group for each measure. Results identified 29 outcome measures that met the eligibility criteria including eight cognition and emotion measures, nine functioning measures, six quality of life measures, and six multidimensional measures.\textsuperscript{37} It revealed that the design of outcome measures has been in line with the traditional split of mental health services between those younger than 18 years and those 18 years and older.\textsuperscript{38}

There were five measures within this review that were shown to have been used across the age range of 12–25 years, although these measures were originally developed and validated for an adult population. The review also found three ROMs used in this age group, the Y-OQ, S × S-SDQ, and SFSS, but all were designed to be used with children and adolescents and had not been validated with young adults older than 18 years.\textsuperscript{37} The reorientation of mental health services means outcome measures specific to the youth transition period spanning adolescence and young adulthood is urgently needed to track outcomes relevant to young people.\textsuperscript{37,38} Additionally, ROM feedback provides another therapeutic tool for supporting this highly vulnerable age group.\textsuperscript{37}
Development of MyLifeTracker (MLT)

headspace National Youth Mental Health Foundation is an Australian innovation in youth mental health services offering mental health, alcohol and other drugs, health care, and vocational services to young people aged 12–25 years through a primary care platform. Established with funding from the Australian government in 2006, by 2018, there will be >100 headspace centers across Australia. The centers prioritize an early intervention approach to youth mental health and are designed to redress the barriers to help-seeking that young people experience, by being no or low cost, youth focused, inclusive, and nonstigmatizing. The centers are supported by an online service, headspace, which facilitates even earlier intervention via online chat, email, and a phone service.

In 2013, headspace centers implemented an innovative data collection system that collects information at each occasion of service for a young person. The data set has multiple purposes, including service monitoring and evaluation and providing therapeutically useful information for clinicians. A particular challenge, however, was determining appropriate mental health outcome measures for use with this age range, as there were none available that targeted those aged 12–25 years or that were appropriate across a diverse range of mental health presentations. In response, headspace developed a measure called “MyLifeTracker” to supplement the other measures being used in the data collection system.

MLT is a brief mental health outcome measure designed for young people aged 12–25 years. Its items measure current self-reported quality of life in the following five different areas of importance to young people: general well-being, day-to-day activities, relationships with friends, relationships with family, and general coping. The measure is intended to be used at every session. It is completed electronically on a tablet or desktop device, and the information is immediately available to the clinician in the visual representation of graphs over time. The measure was developed specifically for use in headspace through a literature review; consultation with clinical staff, clients, and the headspace youth reference group; and pilot testing. The aim was to develop an outcome measure that reflected the important areas of life for young people aged 12–25 years and that could be used for routine monitoring of client progress by being responsive to change.

The current study examines the psychometric properties of MLT to determine its reliability, validity, sensitivity to change, and whether it is appropriate for the entire age range from 12 to 25 years and for both males and females. It was hypothesized that the five items constituting MLT would represent a single factor and that it would show good internal consistency. MLT was expected to display adequate test–retest reliability and show appropriate concurrent validity with standardized mental health outcome measures. Finally, it was hypothesized that MLT would display appropriate sensitivity to change over time.

Participants and methods

Participants

Participants were 75,893 adolescents and young adults between the ages of 12 and 25 years who commenced their first episode of care at a headspace center for a mental health-related issue between July 1, 2015 and March 31, 2017. A total of 25.3% of the participants were in early adolescence (12–14 years), 31.9% of the participants were in mid-adolescence (15–17 years), 28.3% of the participants were in late adolescence (18–21 years), and 14.4% of the participants were young adults (22–25 years). There were more females (58.3%) than males (40.4%), and 1.3% of them reported a nonbinary gender option. Participants were predominantly from major cities (66.0%), followed by inner regional areas (24.5%), outer regional areas (8.4%), and remote areas (1.1%). A total of 8.5% of the participants were Aboriginal and Torres Strait Islander young people. A total of 8.9% of the participants were from culturally and linguistically diverse backgrounds, and 17.6% of the participants identified their sexuality as lesbian, gay, bisexual, other sexualities, or questioning.

At initial presentation, the majority of participants had not been assessed or did not meet threshold diagnostic criteria for a mental disorder; however, primary presenting issue was reported by the client and clinician. As the service is aimed at early intervention, young people who are at-risk or have subclinical presentations are encouraged to present. The initial primary presenting issues include depressive symptoms (29.6%), anxiety symptoms (28.6%), anger issues (7.4%), stress related (5.0%), conflict in home environment (4.6%), difficulty with personal relationships (3.9%), alcohol and other drug use (2.8%), behavioral problems (2.5%), suicidal thoughts or behaviors (2.3%), trauma (2.2%), grief (1.6%), bullying (1.3%), and a range of other mental health symptoms and situational stressors (≤1%).

Procedure

On presentation for each occasion of service at a headspace center, young people were given access to an electronic tablet device or a desktop computer where they were provided with information about the data collection system and consent to answering a series of questions, which varied depending...
on the occasion of service. The time taken to complete the questions varied according to visit as different question sets were presented at different times. The initial questions at first presentation took up to 15 minutes to complete; questions at subsequent visits generally took <5 minutes. Clinicians also completed a series of questions related to each occasion of service after the young person’s session. The data were encrypted and uploaded to a national data warehouse, which is used for monitoring and evaluation. The data set includes demographic characteristics, clinical presentation, and outcome measures.

Ethics approval was obtained through quality assurance processes, comprising initial consideration and approval through the headspace Board research subcommittee. The consent processes were reviewed and endorsed by an independent body, the Australasian Human Research Ethics Consultancy Services. Follow-up data collection processes were approved by Melbourne Health Quality Assurance ethics. The data were collected primarily for service provision and quality improvement purposes, and parental consent was not routinely sought for those younger than 18 years, who were assessed as mature minors unless otherwise indicated. Young people were able to opt out of data collection if they chose, or at the discretion of their parents.

**Measures**

Demographic characteristics comprised age, gender, sexual orientation, Aboriginal and/or Torres Strait Islander background, cultural and linguistic backgrounds, living situation, and current work/study status. Client clinical presentation was determined by self-reported reasons for presentation and clinician-reported presenting issues.

**Outcome measures**

Routine outcome monitoring measure

MLT is a five-item self-report measure to assess current quality of life in areas of importance to young people. It asks young people to self-report how they have been feeling over the last week, or, since their last headspace session, in relation to their “general well-being (emotional, physical, spiritual)”, “day-to-day activities (study, work, leisure, self-care)”, “relationships with friends”, “relationships with family”, and “coping (dealing with life, using your strengths)”. Responses are given on a sliding scale anchored at 0 and 100 with the chosen score visible; the anchors are accompanied by a visual analog of a sad and happy face. Total MLT scores were calculated by averaging across the five items, ranging from 0 to 100, with a higher score indicating a higher quality of life.

**Psychological distress**

The Kessler Psychological Distress Scale 10 (K10) is a 10-item self-report measure of nonspecific psychological distress in the anxiety-depression spectrum. An example item is “In the last 4 weeks, how often did you feel hopeless?” The response format ranged from 1 (none of the time) to 5 (all of the time) for each item. Scores are summed, and the total score ranged from 10 to 50, with higher scores indicating more distress. The K10 has been validated for use in the general Australian population. Cronbach’s $\alpha$ was 0.91 in the present study.

**Life satisfaction**

Life satisfaction is a seven-item self-report measure that was adapted from the Brief Multidimensional Students’ Life Satisfaction Scale from the Peabody Treatment Progress Battery. The questions asked young people to indicate how satisfied they are with different areas of their life, in relation to “family life”, “friendships”, “romantic relationships”, “school/work experience”, “yourself”, “where you live”, and “life overall”. The response format was a 0–10 scale for each item, 0 (worst possible) to 10 (best possible). Total life satisfaction scores were calculated by averaging across the seven items, ranging from 0 to 10, with a higher score indicating more positive life satisfaction. The original measure was validated with young people aged 11–18 years in the USA and had a Cronbach’s $\alpha$ of 0.77. Cronbach’s $\alpha$ was also 0.77 in the present study.

**Well-being**

Three items taken from the 14-item Warwick–Edinburgh Mental Well-Being Scale (WEMWBS) were used to measure self-reported positive mental health and well-being. The three items were selected through consultation with headspace young people and staff. The questions asked young people to rate the frequency of their feelings and thoughts over the last 2 weeks, in relation to “I’ve been interested in new things”, “I’ve been feeling useful”, and “I’ve been feeling good about myself”. Responses were on a 5-point scale from 1 (none of the time) to 5 (all of the time). Total well-being scores were calculated by averaging across the three items, ranging from 1 to 5, with a higher score indicating more positive well-being. WEMWBS has been validated with teenagers in England and Scotland revealing a Cronbach’s $\alpha$ of 0.87. Cronbach’s $\alpha$ was 0.72 in the present study.

**Social and occupational functioning**

The Social and Occupational Functioning Assessment Scale (SOFAS) is a single-item, clinician-report assessment of a
client’s current functioning at the time the evaluation is made, determined irrespective of the severity of the client’s psychological symptoms. Scores range from 1 to 100, with higher scores representing increased functioning across a variety of activities. For example, a score of 1–10 is indicative of a persistent inability to maintain minimal personal hygiene; a score of 41–50 is indicative of serious impairment in social, occupational, or school functioning; and a score of 91–100 is indicative of superior functioning in a wide range of activities. The SOFAS is derived from the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) Global Assessment of Functioning Scale (Axis V), which has shown good internal consistency with the Cronbach’s α of 0.80.43

Data analyses
First, we conducted an exploratory factor analysis (EFA) of MLT items using maximum likelihood factoring. A confirmatory factor analysis (CFA) was then conducted to replicate the factor structure of MLT. The baseline data set was randomly split into two, with the EFA performed on one half and the CFA performed on the other. As the Chi-square test is highly sensitive to sample size, the following fit indices were used for the CFA to assess the model fit: root mean square error of approximation (RMSEA < 0.06, with the lower bound of its 90% CI < 0.05 to indicate “close” fit), the comparative fit index (CFI > 0.95), the Tucker–Lewis Index (TLI > 0.95), and the Standardized Root Mean Square Residual (SRMR < 0.08).44 Additionally, a multi-group CFA was performed to test measurement invariance across age and gender groups.45 The following indices have been recommended: CFI>0.95 for configural invariance and ACFI ≤ 0.01 for metric invariance.46

Internal consistency of MLT using the first visit data was evaluated using the Cronbach’s α coefficient. Pearson’s correlation analysis was conducted to examine the test–retest reliability of MLT scores across the first and second headspace visits that were within a 2-week time period. Internal consistency and test–retest reliability were further examined as a function of age group and gender (male/female). Concurrent validity was examined using a Pearson’s correlation analysis, which explored the associations of MLT scores at participants’ initial session with the standardized outcome measures of psychological distress, life satisfaction, well-being, and social and occupational functioning. Concurrent validity was further examined as a function of age group and gender.

Sensitivity to change was examined using a number of methods including effect sizes,47 Reliable Change Index (RCI),48 and growth curve modeling.49 Effect sizes were estimated for participants’ average MLT change scores between the first and last sessions. RCIs were calculated between participants’ first and last MLT scores, and percentages of participants who improved, did not change, or deteriorated were reported. RCIs were also calculated for K10 scores to allow comparisons with MLT. RCI is an index that was developed to detect the minimum reliable amount of change in scores while accounting for measurement error.48 Effect sizes and RCIs were further examined as a function of age group and gender.

Growth curve modeling estimated average rates of change in MLT scores across participants’ episode of care. This approach was utilized as it can estimate a mean intercept and mean slope of the pooled sample trajectory (within-person patterns).49 Maximum likelihood estimation procedures were used. Weeks in treatment were used over session number as time between sessions varied among participants, and this approach has been recommended in the literature exploring youth psychotherapy change.50 Only cases below the 90th percentile (26 weeks) of treatment length were used to avoid effects of outliers. Inspection of MLT data suggested a cubic pattern rather than a linear pattern over time (Bayesian Information Criterion [BIC]linear = 1892911, BICcubic = 1890673, and BICcubic = 2238, where smaller BICs indicate a better model fit). Growth curve modeling was further examined as a function of age group and gender. All analyses were conducted using SPSS Version 23 and Amos Version 23.

Results
Descriptives
Table 1 provides descriptive statistics for MLT scores and the standard outcome measures at first headspace visit. It reveals that females generally reported greater psychological distress, lower life satisfaction, and lower well-being than males (P < 0.001, $\eta_p^2$ = 0.04, 0.01, and 0.02, respectively). In addition, young adults generally reported greater psychological distress, lower life satisfaction, and lower well-being than adolescents (P < 0.001, $\eta_p^2$ = 0.07, 0.06, and 0.03, respectively). These trends across age and gender groups were also seen with the MLT scores (P < 0.001, $\eta_p^2$ = 0.06 and 0.03, respectively). Clinician reports of participants’ social and occupational functioning appeared constant across gender and age group (P < 0.001, $\eta_p^2$ = 0.004 and 0.001, respectively).

Exploratory factor analysis
The eigenvalues generated by the EFA were 3.17, 0.67, 0.64, 0.32, and 0.21. Standardized factor loadings for the five MLT items were uniformly strong and positive (range 0.53–0.90).
The one factor solution explained 56% of the variance. The eigenvalues and factor loadings supported a unidimensional solution for MLT.

**Confirmatory factor analysis**
Fit indices obtained from the CFA indicated an excellent level-of-fit of the data to the single-factor model: RMSEA=0.049, 90% CI (0.045, 0.053), CFI=0.995, TLI=0.990, and SRMR=0.015. Measurement invariance tests across age and gender groups indicated that MLT displayed configural invariance (CFI=0.990) and metric invariance (ΔCFI=0.006).

**Internal consistency**
MLT showed good internal consistency with Cronbach’s α=0.84. Internal consistency was further examined as a function of age group and gender, and Cronbach’s α ranged from 0.79 to 0.86.

**Test–retest reliability**
Due to variability of time between participants’ initial and second sessions, only a test–retest period within a 2-week time frame was applied (n=18890). MLT test–retest reliability showed a coefficient of 0.78. Test–retest reliability was further examined as a function of participant age group and gender. Coefficients were highly similar, ranging from 0.74 to 0.78 across the age group and gender combinations.

**Concurrent validity**
Table 2 provides the correlation coefficients for the relationships of MLT with the standardized outcome measures. MLT was strongly negatively correlated with psychological distress and strongly positively correlated with life satisfaction and well-being. It was also positively correlated with social and occupational functioning, but these associations were weak in strength. The correlation coefficients between MLT and the standardized outcome measures were consistent in strength across age and gender groups.

**Sensitivity to change**
The number of sessions attended by participants ranged from 1 to 30 sessions, with an average attendance of 4.16 sessions (SD=4.37). Change was determined by comparing the first and last session scores. Overall, participants demonstrated

### Table 1: Descriptive statistics for MyLifeTracker and standardized outcome measures at first headspace visit, by age group and gender

<table>
<thead>
<tr>
<th>Outcome measures</th>
<th>12–14 years, mean (SD)</th>
<th>15–17 years, mean (SD)</th>
<th>18–21 years, mean (SD)</th>
<th>22–25 years, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>n=6243</td>
<td>n=7772</td>
<td>n=7451</td>
<td>n=3903</td>
</tr>
<tr>
<td>MyLifeTracker</td>
<td>mean (age)=13.07</td>
<td>mean (age)=16.00</td>
<td>mean (age)=19.37</td>
<td>mean (age)=23.09</td>
</tr>
<tr>
<td>Psychological distress</td>
<td>61.88 (21.14)</td>
<td>52.70 (20.17)</td>
<td>45.78 (19.51)</td>
<td>44.35 (19.33)</td>
</tr>
<tr>
<td>Life satisfaction</td>
<td>21.63 (8.19)</td>
<td>25.79 (8.75)</td>
<td>29.05 (8.63)</td>
<td>29.51 (8.60)</td>
</tr>
<tr>
<td>Well-being</td>
<td>6.54 (1.83)</td>
<td>5.80 (1.79)</td>
<td>5.09 (1.84)</td>
<td>4.85 (1.86)</td>
</tr>
<tr>
<td>Social and occupational functioning</td>
<td>63.50 (17.09)</td>
<td>62.87 (16.37)</td>
<td>60.77 (16.82)</td>
<td>61.06 (17.06)</td>
</tr>
<tr>
<td>Females</td>
<td>n=9818</td>
<td>n=12,474</td>
<td>n=10,000</td>
<td>n=4786</td>
</tr>
<tr>
<td>MyLifeTracker</td>
<td>mean (age)=13.21</td>
<td>mean (age)=15.93</td>
<td>mean (age)=19.36</td>
<td>mean (age)=23.03</td>
</tr>
<tr>
<td>Psychological distress</td>
<td>49.21 (20.51)</td>
<td>43.45 (18.09)</td>
<td>41.05 (17.80)</td>
<td>41.73 (18.48)</td>
</tr>
<tr>
<td>Life satisfaction</td>
<td>27.08 (9.11)</td>
<td>30.36 (8.49)</td>
<td>31.54 (8.26)</td>
<td>30.96 (8.24)</td>
</tr>
<tr>
<td>Well-being</td>
<td>5.69 (1.82)</td>
<td>5.29 (1.69)</td>
<td>5.04 (1.72)</td>
<td>4.98 (1.75)</td>
</tr>
<tr>
<td>Social and occupational functioning</td>
<td>63.70 (16.25)</td>
<td>62.96 (16.06)</td>
<td>62.01 (16.59)</td>
<td>62.37 (17.11)</td>
</tr>
</tbody>
</table>

### Table 2: Pearson correlation coefficients between MyLifeTracker and standardized outcome measures, by age group and gender

<table>
<thead>
<tr>
<th></th>
<th>12–14 years</th>
<th>15–17 years</th>
<th>18–21 years</th>
<th>22–25 years</th>
<th>All ages</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD, LS, WB, SOF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MyLifeTracker</td>
<td>-0.61</td>
<td>-0.65</td>
<td>-0.66</td>
<td>-0.67</td>
<td>-0.68</td>
</tr>
<tr>
<td>MyLifeTracker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MyLifeTracker</td>
<td>-0.66</td>
<td>-0.64</td>
<td>-0.64</td>
<td>-0.65</td>
<td>-0.66</td>
</tr>
</tbody>
</table>

**Note:** All P<0.05.

**Abbreviations:** PD, psychological distress; LS, life satisfaction; WB, well-being; SOF, social and occupational functioning.
change with a moderate effect size ($d=0.52$). Effect sizes were similar across the age group and gender combinations, with Cohen’s $d$ ranging from 0.47 to 0.58.

RCI was estimated as an 18.27-point change for MLT, and Table 3 provides the percentages of young people showing reliable change across age group and gender. The proportions of reliable change for the K10 are also provided in Table 4 for comparison with MLT. RCI was estimated as a 7-point change for the K10 based on previous research.36 MLT showed more participants reliably improving compared to the K10 ($–1.5\%$) and fewer participants reliably deteriorating than the K10 ($+2.5\%$). Therefore, MLT detected more improvement and less deterioration among the participants when compared with the K10.

Change in MLT followed a cubic pattern, which had a significant contribution in the model ($P<0.001$). The initial status of MLT was $–46.0$ ($β=46.53$, standard error $[SE]=0.09$, $P<0.001$). The positive effect of linear growth ($β=1.48$, $SE=0.02$, $P<0.001$) suggested that the MLT increased at the beginning; the negative effect of the quadratic growth ($β=–0.08$, $SE=0.003$, $P<0.001$) indicated a deceleration in the rate of change; and the positive effect of cubic growth ($β=0.002$, $SE=0.001$, $P<0.001$) revealed that such deceleration gradually diminished over time. As shown in Figure 1, the average rate of change during early therapy was relatively steep but began to flatten out between weeks 10 and 20; then, between weeks 21 and 26, there was another slight upward trend.

Age and gender groups were examined as a time-invariant covariate to explore any group differences in change over time. Although the initial scores for MLT were significantly different across age and gender combinations ($β=–2.36$, $SE=0.05$, $P<0.001$), as shown in Table 1, the shape of the change did not differ significantly across age and gender groups ($P>0.05$).

### Table 3

Proportion of young people showing reliable change for MyLifeTracker between first and last session ratings, by age group and gender

<table>
<thead>
<tr>
<th>Change category (%)</th>
<th>12–14 years</th>
<th>15–17 years</th>
<th>18–21 years</th>
<th>22–25 years</th>
<th>All participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement</td>
<td>29.0</td>
<td>28.4</td>
<td>28.0</td>
<td>30.2</td>
<td>28.2</td>
</tr>
<tr>
<td>No change</td>
<td>65.4</td>
<td>66.7</td>
<td>67.3</td>
<td>65.8</td>
<td>66.8</td>
</tr>
<tr>
<td>Deterioration</td>
<td>5.6</td>
<td>4.9</td>
<td>4.7</td>
<td>4.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

### Table 4

Proportion of young people showing reliable change for K10 between first and last session ratings, by age group and gender

<table>
<thead>
<tr>
<th>Change category (%)</th>
<th>12–14 years</th>
<th>15–17 years</th>
<th>18–21 years</th>
<th>22–25 years</th>
<th>All participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement</td>
<td>23.7</td>
<td>25.5</td>
<td>27.9</td>
<td>29.3</td>
<td>26.7</td>
</tr>
<tr>
<td>No change</td>
<td>68.2</td>
<td>66.5</td>
<td>65.8</td>
<td>64.8</td>
<td>65.8</td>
</tr>
<tr>
<td>Deterioration</td>
<td>8.1</td>
<td>8.0</td>
<td>6.3</td>
<td>5.8</td>
<td>7.5</td>
</tr>
</tbody>
</table>

**Abbreviation:** K10, Kessler Psychological Distress Scale 10.

Discussion

The current study aimed to examine the psychometric properties of MLT to determine its reliability, validity, sensitivity to change, and whether it is appropriate for the entire age range from 12 to 25 years and for both males and females.

A unidimensional factor structure was determined using EFA and, then, confirmed through a CFA. This provides support for clinicians and services to use MLT as a single total score to track mental health outcomes. Importantly, the measure was shown to display measurement invariance across age and gender combinations. The reliability of MLT was demonstrated through good internal consistency and robust test–retest reliability.

Evidence for the concurrent validity of MLT was found, as correlations between MLT and standardized outcome measures were as hypothesized. Specifically, MLT was strongly related to psychological distress, life satisfaction, and well-being. Weak correlation coefficients between the client-reported MLT and the clinician-rated SOFAS argue in favor of the necessity of implementing routine client-reported outcome measures rather than relying solely on clinician subjective judgment of treatment progress.8,11,12 Past research has shown significant differences between self-report versus clinician-reported ratings of similar measures.51,52 Similar correlation coefficients were evident.
for various age groups and for males and females, further demonstrating concurrent validity across age and gender for MLT.

The sensitivity to change of MLT was shown via significant improvements that were observed for males and females in all age groups. Importantly, these changes were moderate in effect size. RCI proportions across MLT and the K10 were generally comparable, but MLT showed higher proportions of participants improving and lower proportions of participants deteriorating when compared with the K10. It should be noted, however, that some participants would not be able to reliably change as their initial score may be too close to the upper or lower limits of the measure (eg, if a participant’s initial score on MLT was 82, they could not reliably improve by the required 18.27 points as the upper limit of the measure is 100).

MLT RCI percentages and effect sizes are comparable to those reported from other naturalistic youth mental health settings. A study in the Netherlands explored naturalistic change in a youth mental health clinic targeting young adults aged 16–23 years attending an average of 7.5 sessions. The results showed an overall effect size of 0.55 and 19% of clients reliably improving when rated on the Global Assessment of Functioning scale.53 A study exploring change in parent-reported SDQ in England CAMHS showed an overall effect size of \( d=0.40 \), and reliable improvement was seen among 16.5% of clients after a 4- to 8-month period.54 Australian data from acute inpatient CAMHS from 2006 to 2007, using the Health of the Nation Outcome Scales for Children and Adolescents, showed 23.8% reliable improvement and 1.8% reliable deterioration from admission to discharge.55

The growth curve pattern demonstrated by MLT showed a steep increase during the early stages of therapy, subsequent flattening out, and then a slight increase again later in therapy. There were similar rates of change across age and gender groups. This cubic pattern of change has been observed in other research with children and adolescents using the parent-reported Y-OQ.50 Similarly, this pattern of

Figure 1 Average rate of change in MLT across weeks in treatment.
Abbreviation: MLT, MyLifeTracker.
change was seen with clients from a university counseling center using the OQ-45, which is a self-report measure for adults. A theory for this pattern of change is the “good enough level” model, whereby clients are said to change at different rates and those that have reached sufficient change terminate treatment, and therefore, a plateau is seen within the growth curve. Interestingly, there is some evidence that gains, particularly sudden gains, occur earlier in the treatment for younger clients. For example, Gaynor et al. found that in children and adolescents, 85% of sudden gains occurred before session five and all sudden gains by the 10th session; whereas for an adult sample, only half of sudden gains had occurred by session five. As such, rates of change shown by MLT is comparable to other well-validated measures in similar youth-based settings, however, due to the naturalistic clinical setting in the current sample, MLT could not be compared with a control group.

This study offers evidence for the validity of MLT as a brief and practical measure that can be used in routine outcome monitoring with adolescents and young adults. Information is quickly and easily obtained from young people and able to be displayed graphically for clinicians to track treatment progress. MLT fills a gap in mental health outcome measures for young people aged 12–25 years. The measure has been specifically designed for and now validated with this age group, which is preferable to using measures designed for adult populations. MLT targets valued parts of a young person’s life including relationships, coping, and well-being, rather than focusing on traditional measures of diagnostic symptoms and functioning. The shift of mental health service provision to the transition period of adolescence and early adulthood means outcome measures for this age range are vital, and MLT has been shown to be psychometrically sound across this entire age range. The measure enables clinicians working with young people to gain a quick regular snapshot of overall client progress and provides a valid measure to use to assess service effectiveness.

Furthermore, each item on MLT targets current quality of life in areas of importance to young people themselves, including general well-being, day-to-day activities, relationships with friends, relationships with family, and coping. These areas were developed through extensive consultation with headscape clients and their clinicians and shown to be meaningful to young people and clinically useful to clinicians. While the items are all tapping into a single dimension, the level of change within each item can identify areas where therapy is being effective or not – for example, relationships and coping skills. If outcomes are routinely monitored by clinicians and used as a therapeutic tool with clients, this can improve outcomes, specifically for identifying clients who are deteriorating or not improving. Such monitoring provides clinicians with another source of information on a session-by-session basis to help with treatment planning.

However, the current results should be considered in light of several limitations. First, the study was conducted in Australia within an early intervention youth mental health setting and the psychometric properties of MLT may not be generalizable to other settings, such as public mental health services, or other age groups, such as children younger than 12 years or older adults. Further testing is required to support it being used more broadly. Another limitation is that the current study has not provided norms or clinically significant change values. This would provide a general benchmark for clinicians to compare against when providing treatment to young people and enhance the practical utility of the measure to aid in its interpretation.

Further research should explore clinicians’ and young people’s views of MLT to confirm that it measures meaningful changes. A qualitative approach has been suggested to better understand this developmentally sensitive construct of therapeutic change. Furthermore, future research should consider the clinical utility of MLT as a feedback system in routine outcome monitoring to determine how clinicians and clients can use the measure to discuss outcomes and areas of need for targeted intervention.

The push for ROM should not only be motivated by a need for increased quality assurance and effectiveness but also provide clinically useful and meaningful tools for clinicians and clients to enhance the therapeutic process. MLT provides a psychometrically sound mental health outcome measure for young people aged 12–25 years. The measure is brief and easy to use with a young person. It can be incorporated into an electronic system that routinely tracks session-by-session change and produce time-series charts for the ease of use and interpretation. MLT taps into areas that are meaningful to young people and provides an additional support to clinicians and clients during therapy.

**Author contributions**
BK completed the literature review, analyzed the data, and drafted the article. DJR supervised the design, reviewed the results, and revised the article. NRT supervised the design and reviewed the results. All authors contributed toward data analysis, drafting and critically revising the paper and agree to be accountable for all aspects of the work.
Disclosure

All authors are either employed by headspace National or a headspace center. The authors report no other conflicts of interest in this work.

References

Chapter V: A Routine Outcome Measure for Youth Mental Health: Clinically Interpreting MyLifeTracker

Chapter V presents the third paper prepared during the PhD candidature. The paper aimed to provide statistically derived benchmarks for MLT for clinical use in youth mental health settings. Paper Three has been peer reviewed and was published in *Early Intervention in Psychiatry* in July 2020. *Early Intervention in Psychiatry* is a Q1 journal in the Psychiatry and Mental Health category and Psychiatric Mental Health category, with an h-index of 42 and an impact factor of 2.257. This paper has been cited in one journal article since publication. The format of the paper is in accordance with the publication guidelines for this journal.
Declaration of Co-Authored Publications

For use in theses which include co-authored publications. This declaration must be completed for each co-authored publication and to be placed at the start of the thesis chapter in which the publication appears, or as a preface to the thesis.

Declaration for Thesis Chapter V

DECLARATION BY CANDIDATE

In the case of Chapter V, the nature and extent of my contribution to the work was the following:

<table>
<thead>
<tr>
<th>Nature of Contribution</th>
<th>Extent of Contributions (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptualisation and design of the study, review of the literature, data analysis, and write up of the article.</td>
<td>75%</td>
</tr>
</tbody>
</table>

The following co-authors contributed to the work:

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
<th>Contributor is also a UC student (Yes/No)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debra Rickwood</td>
<td>Conceptualisation and design of the study, review of the results, and editing the article.</td>
<td>No</td>
</tr>
</tbody>
</table>

Candidate's Signature

Date: 18/10/2020

DECLARATION BY CO-AUTHORS

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
(4) there are no other authors of the publication according to these criteria;
(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Location(s): University of Canberra

Signatures

Date: 19/10/2020

1 of 1
A routine outcome measure for youth mental health: Clinically interpreting MyLifeTracker

Benjamin Kwan1 | Debra J. Rickwood1,2

1Faculty of Health, University of Canberra, Bruce, Australian Capital Territory, Australia
2headspace National Youth Mental Health Foundation, Melbourne, Victoria, Australia

Abstract
Aim: MyLifeTracker is a session-by-session mental health outcome measure for young people aged 12 to 25 years. The aim of this study was to determine clinically significant change indexes for this measure that would identify developmentally appropriate thresholds. The study also aimed to determine expected change trajectories to enable clinicians to compare a client’s progress against average rates of change.

Methods: Participants comprised young people aged 12 to 25 years from both a clinical and a community sample from Australia. The clinical sample was 63 840 young people that attended a headspace centre. The non-clinical group was an Australian representative community sample of 4034 young people.

Results: Clinically significant change indexes were developed for MyLifeTracker specific for age and gender groups by comparing clinical and non-clinical samples. Males and young people aged 12 to 14 years needed to reach higher scores to achieve clinically significant change compared to females and other age groups, respectively. MyLifeTracker expected change trajectories followed a cubic pattern for those with lower baseline scores of 0 to 50, whereas those with baseline scores of 51 and above had varying patterns. For those with lower baseline scores, expected change trajectories showed that stronger change was evident early in treatment, which then tapered off before accelerating again later in treatment.

Conclusions: The development of MyLifeTracker benchmarks allows the measure to be used for Feedback Informed Treatment by supporting treatment planning and decision-making. This information can help clinicians to identify clients who are not on track or deteriorating and identify when clients are improving.

KEYWORDS
clinically significant change, expected change, MyLifeTracker, routine outcome measure, youth mental health

1INTRODUCTION

MyLifeTracker (MLT) is a recently developed mental health outcome measure for routine monitoring specifically targeted for young people aged 12 to 25 years. It was co-designed with both young people and youth mental health clinicians to assess meaningful outcomes in the domains of general wellbeing, daily functioning, relationships with friends, relationships with family and coping. MLT has shown evidence...
of a single factor structure, although the five items were also designed to be clinically useful individually. Overall, MLT measures the quality of life, with higher scores indicating higher levels of quality of life. It has been validated against measures of psychological distress, quality of life and wellbeing, and demonstrates appropriate reliability and sensitivity to change (Kwan, Rickwood, & Telford, 2018). MLT is currently implemented into an electronic data system used by the 110 headspace youth mental health services implemented across Australia (Rickwood et al., 2018). This electronic data system provides information to clinicians that are collected from clients prior to every visit and displayed to the clinician in the form of a graph over time of MLT scores. This reveals change over time that can be used by the clinician to ascertain treatment progress and can also be shown to clients during their session via a computer or tablet device.

MLT was developed to fill a measurement gap in youth mental health. Historically, outcome measures have been designed that reflect the traditional mental health service demarcation between the child and adolescent services, for those aged less than 18 years, and adult services, for those aged 18 years and above (Kwan & Rickwood, 2015). The growing implementation of youth mental health services internationally, which span the age range of 12 to 25 years, necessitates new measures (McGorry, Bates, & Birchwood, 2013). MLT was designed to be suitable for session-by-session use by being very brief and comprising only five items. An essential next step for the development of MLT is the identification of age and gender norms, which improves the interpretability of such measures (Centofanti et al., 2018). This information is particularly important in youth mental health because adolescence and early adulthood are periods of rapid social, emotional and physical development when age and gender differences are pronounced (Donald, Rickwood, & Carey, 2014; Rickwood et al., 2015).

The development of session-by-session measures for routine outcome monitoring supports Feedback-Informed Treatment (FIT) approaches, whereby a clinician receives quantitative feedback on a client’s progress to use in-session and help guide treatment planning (Bickman, 2008). FIT requires a measurement system that is easily completed by the client and allows instant feedback to the clinician (Hall et al., 2014). This provides clinicians with regular up-to-date snapshots of a client’s mental health status and shows any changes since past sessions (Lutz, De Jong, & Rubel, 2015). Clinicians are then able to monitor if clients are progressing or deteriorating between sessions, and adjust treatment planning accordingly (Boswell, Kraus, Miller, & Lambert, 2015).

Such an approach can also allow clinicians to bring the measures into sessions and feedback progress to clients, which can be a powerful way to promote shared decision making (Reese, Norsworthy, & Rowlands, 2009). FIT has been shown to improve communication between client and clinician, increase the accuracy of diagnosis, enable quicker adjustments to treatment planning when required, provide stronger outcome effects and improve the efficiency of treatment (Bickman, Kelley, Breda, de Andrade, & Riemer, 2011; Carlier et al., 2012; Janse, De Jong, Van Dijk, Hutschemaekers, & Verbraak, 2017).

A valuable metric for clinicians to use in FIT is change scores, such as clinically significant change indexes, expected change trajectories and early warning signals. These can be calculated from session-by-session measures to provide evidence-based benchmarks for FIT and routine outcome monitoring systems. A clinically significant change is conceptualized as the process of a client starting treatment in the dysfunctional (clinical) population and leaving treatment no longer in this population (Jacobson, Follette, & Revenstorf, 1984). It is operationalized as a change in a client’s outcome measure score showing that they are statistically more likely to be drawn from the functional distribution, having moved out of the dysfunctional distribution during treatment (Jacobson & Truax, 1991). When the dysfunctional and functional populations are identified, clinically significant change indexes can be calculated by finding the value where the two populations intersect. Reliable change can also be determined, which takes into account the reliability of the outcome measure, ensuring that change is not due to measurement error.

Change can be then categorized into four stages: Deterioration—when a client has reliably worsened; Unchanged—when no reliable change has occurred; Improvement—when a client has made a reliable positive change but still remains in the dysfunctional population and Recovered—when a client reliably improves and moves into the functional population (Jacobson & Truax, 1991).

A criticism of clinically significant change is that it can be an overly stringent measure of change, being based on diagnostic cut-offs. In naturalistic clinical settings, some clients may not be able to reach this threshold because they initially present in the functional population range (Wise, 2004). Other methods of monitoring change have been recommended; specifically, the use of growth curve modeling, which shows expected rates of change (Donald & Carey, 2017). This approach estimates a mean starting point (intercept) and average rates of change (slope) of the pooled sample trajectory; that is, within-person expected change patterns (Singer & Willett, 2003). The method is particularly useful for exploring client change in naturalistic therapy settings as it can deal with data that are time-unstructured and unbalanced. This provides clinicians with an expected change trajectory, which can be compared with an individual client’s trajectory to determine whether the client is within or outside expected rates of change, potentially indicating the cause for concern (Finch, Lambert, & Schaalje, 2001).

Research has increasingly focussed on detecting clients who are at risk of deterioration using early warning systems that are derived from expected change trajectories (Finch et al., 2001). An early warning is evident when a client’s score drops below an identified threshold. It is recommended that these early warning signals be derived from the bottom-end percentage of the targeted population and the proportion of clients who reliably deteriorate in that population (Finch et al., 2001; Warren, Nelson, Mondragon, Baldwin, & Burlingame, 2010). An essential aspect of early warning signals is the ability to accurately predict clients who are responding poorly to treatment or are not on track (NOT) before therapy is terminated (Boswell et al., 2015). Some studies have evaluated the efficacy of these signals of deterioration, alerting clinicians to clients that are falling into the bottom 10% to 20%, demonstrating detection accuracy rates of 85% to 100% when used with adult clients (Lambert et al., 2002). Lower detection accuracy rates of
69% to 88% are seen when early warning signals are used with children and adolescents, which has been justified by the higher proportions of treatment failure when compared to adult clients (Cannon, Warren, Nelson, & Burlingame, 2010; Nelson, Warren, Gleave, & Burlingame, 2013; Warren, Nelson, & Burlingame, 2009).

Therapeutic deterioration is evident in up to 10% of adult clients (Lilienfeld, 2007; Murphy, Rashleigh, & Timulak, 2012), but much higher at 21%, for clients in youth psychotherapy settings (Warren et al., 2009). High dropout rates are another major concern in youth mental health settings, and dropout has been shown to be partly due to clinician and therapeutic factors that may be responsive to feedback (de Haan, Boon, de Jong, Hoeve, & Vermeiren, 2013). Early warning alerts have been shown to reduce deterioration from 20.1% in treatment as usual to 5.5% in feedback conditions for adult clients (Shimokawa, Lambert, & Smart, 2010). Feedback was also shown to double the proportion of clients with clinically significant improvement in NOT clients. Feedback to clinicians alone, and to both clinician and client, has been shown to significantly positively increase the rates of change in short-term adult NOT clients (De Jong et al., 2014).

FIT approaches are increasingly being advocated because clinicians have been shown to have low accuracy rates of predicting client deterioration during therapy when using their judgement alone (Hannan et al., 2005; Hatfield, McCullough, Frantz, & Krieger, 2010). It is proposed that clinicians have a self-assessment bias which serves to maintain a positive self-image (Parker & Waller, 2015). For example, Walfish, McAlister, O’Donnell, and Lambert (2012) explored clinicians’ ratings of their own clinical skills and client outcomes, showing that they rated their skills on average at the 80th percentile and that all clinicians rated themselves above the 50th percentile. In addition, clinicians on average believed that 77% of their clients improved as a result of their therapeutic intervention, which is well above the one-third proportion of clients shown to improve in most naturalistic settings (Walfish et al., 2012). Deliberate practice, incorporating FIT with evidence-based benchmarks, could be very effective at reducing this self-assessment bias amongst clinicians (Chow et al., 2015; Goodyear, Wampold, Tracey, & Lichtenberg, 2017; Macdonald & Mellor-Clark, 2015). Despite the potential clinical utility, however, clinicians have been shown to have limited knowledge about the use of routine outcome measures in predicting client deterioration (Bystedt, Rozental, Andersson, Boettker, & Carlbring, 2014).

The current study investigated the implementation of routine outcome measurement and clinician feedback within youth mental health services, using the MLT measure. We aimed to determine MLT clinically significant change indexes that would identify developmentally appropriate thresholds for different age and gender groups. It was anticipated that there would be different clinically significant change indexes across the developmental period between 12 and 25 years and between males and females, due to the major changes that take place during adolescence and early adulthood and the marked gender differences in mental health status between males and females (eg, females displaying higher levels of psychological distress) (Brann, Lethbridge, & Mildred, 2018; Centofanti et al., 2018; Kwan et al., 2018). Identifying these developmental patterns would allow clinicians to provide more tailored client care. To do this, scores for a clinical population group were compared with data from a nationally representative community sample to determine appropriate change indexes. It was hypothesized that the non-clinical group would have higher MLT scores compared with the clinical group, that males would have higher MLT scores than females, and that the younger adolescents would have higher MLT scores than those who were older (Kwan et al., 2018). We also aimed to determine expected change trajectories and early warning signals for MLT to provide benchmarks to help clinicians identify if a client is showing expected change over time in treatment, or whether the client is deteriorating. Lastly, we provide examples of how clinicians can use the statistically derived benchmarks for MLT in their clinical practice.

2 METHODS

2.1 Participants

Participants comprised both a clinical and a nationally representative community sample. The clinical sample was 63,840 adolescents and young adults between the age of 12 and 25 years who commenced the first episode of care at a headspace centre. headspace is the Australian Government’s National Youth Mental Health Foundation, which was initiated in 2006 to provide early intervention in youth mental health. headspace centres offer services responding to mental health, alcohol and other drugs, general health and vocational concerns for young people (Rickwood et al., 2015). This sample consisted of 40.4% males and 59.6% females, in the following age ranges: 12 to 14 years (24.1%), 15 to 17 years (32.0%), 18 to 21 years (29.1%) and 22 to 25 years (14.8%).

The non-clinical group was a nationally representative community sample that consisted of 4034 young people aged 12 to 25 years from across Australia. The sampling was stratified to provide a near-even split between males (49.1%) and females (50.9%), and across age groups: 12 to 14 years (24.7%), 15 to 17 years (24.7%), 18 to 21 years (25.0%) and 22 to 25 years (25.6%).

2.2 Procedure

The clinical group commenced their first episode of care at a headspace centre between July 1, 2015 and March 31, 2017. During this period, data were available for 101 headspace centres across Australia. headspace centres routinely collect a minimum dataset comprising data from young people and their service providers at every occasion of service. The dataset includes demographic characteristics, clinical presentation and treatment outcome measures. Young people can present for a wide range of reasons to headspace centres (Rickwood et al., 2015), but only those who were deemed by their clinician to be at one of the following stages of mental illness were included in the current analyses: mild to moderate general symptoms; sub-threshold diagnosis; threshold diagnosis; periods of remission or ongoing severe symptoms.
The data from headspace centres are encrypted and uploaded to a national datawarehouse, which is used for research, monitoring and evaluation. Ethics approval was obtained through quality assurance processes, comprising initial consideration and approval through the headspace board research sub-committee. The consent processes have been reviewed and endorsed by an independent body, the Australasian Human Research Ethics Consultancy Services.

The non-clinical group was recruited between July and September 2018. A research consultancy agency was commissioned by headspace to undertake a national computer-assisted telephone interview of young people aged 12 to 25 years from across Australia. A quota sampling procedure was used to ensure equal numbers by gender and age group. The sample was recruited through random digit dialling (RDD; randomly generating Australian mobile phone and landline numbers). Ethics approval was obtained from Bellberry Limited Human Research Ethics Committee.

2.3 Measures

Both the headspace minimum dataset and headspace nationally representative community survey include a large number of demographic, clinical and outcome measures. For the current study, only the demographic characteristics of gender (male, female, other), age group (12-14, 15-17, 18-21 and 22-25 years), and the MLT routine outcome monitoring measure were used.

2.3.1 Routine outcome monitoring measure

MLT (Kwan et al., 2018) is a five-item self-report measure used to assess the current quality of life in areas of importance to young people. It asks young people to indicate how they have been feeling over the last week in relation to their: "general wellbeing (emotional, physical, spiritual)", "day-to-day activities (study, work, leisure, self-care)", "relationships with friends", "relationships with family" and "coping (dealing with life, using your strengths)". Responses are given on a sliding scale anchored at 0 and 100 with the chosen score visible, accompanied by a visual analogue of a sad and happy face as anchors. Total MLT scores were calculated by averaging across the five items, ranging from 0 to 100, with a higher score indicating a higher quality of life. In the present study, internal consistency was high, with the Cronbach’s α = .83 in the clinical group and .88 in the non-clinical group. The original MLT study reported a Cronbach’s α of .84, which ranged from .79 to .86 across age groups and gender (Kwan et al., 2018).

2.4 Data analyses

SPSS V21 was used for all analyses. First, descriptive statistics for MLT were calculated and a factorial between groups analysis of variance (ANOVA) was conducted to evaluate the differences in MLT scores across population groups (clinical, non-clinical), gender (male, female) and age groups (12-14, 15-17, 18-21, 22-25 years). Games-Howell post-hoc tests were conducted to address unequal variances and sample sizes. Due to the large sample size, a significant change was reported as partial η² > .001 and d ≥ .02.

Clinically significant change indexes were calculated using data from the clinical and non-clinical samples for each age group and gender (male and female; there were too few participants reporting non-binary gender in the non-clinical sample to create a third gender group) combinations. Results from the original MLT study revealed differences in baseline MLT scores across age and gender groups (Kwan et al., 2018). The formula proposed by Jacobson and Truax (1991) was used to calculate clinically significant change indexes when both clinical and non-clinical groups are available but have unequal variances (p. 13).

Expected change trajectories were determined for the clinical group using growth curve modelling (Singer & Willett, 2003), which estimated average rates of change in MLT composite scores across participants’ episodes of care. This approach was utilized as it provides fixed effects that estimate a mean slope of the pooled sample trajectory (within-person patterns). Maximum likelihood estimation procedures were used. Weeks in treatment were used over session number as the time variable because this has been recommended in the past literature exploring youth psychotherapy change (Warren et al., 2010) and provided a better model fit based on Bayesian Information Criterion (BIC).

Expected change trajectories were calculated for decile groups dependent on MLT baseline scores; that is, 0 to 10, 11 to 20, etc. A precedence has been set for this method by past research exploring change trajectories, which show differing rates of change dependent on baseline severity on outcome measures (Finch et al., 2001; Lambert et al., 2002). Only data from participants attending more than one session and with treatment length up to 26 weeks were used to avoid extreme outliers in terms of treatment length. Two early warning signals were calculated based on the baseline MLT score and expected change trajectory; one SD below the expected change trajectory and reliable deterioration based on the baseline MLT score.

3 RESULTS

3.1 Clinically significant change indexes

Table 1 provides the descriptives for MLT scores for the clinical and non-clinical groups, and the calculated clinically significant change indexes for MLT across age groups and gender. The ANOVA revealed no significant interactions (partial η² ≤ .001) and only significant main effects. MLT scores were significantly higher in the non-clinical group compared to the clinical group (partial η² = .149); and for males compared with females (partial η² = .005). MLT scores differed significantly by age group (partial η² = .013), and post-hoc analyses revealed that scores for those aged 12 to 14 years were significantly higher than all other age groups (15-17 years (d = .25), 18-21 years (d = .36) and 22-25 years (d = .24)), which did not differ significantly from each other (d < .20).
Females showed a lower threshold to achieve clinically significant change when compared to males across all age groups. Within gender, for both male and female participants, those aged 18 to 21 years showed the lowest threshold for clinically significant change and those aged 12 to 14 years showed the highest threshold when compared to the other age groups. Across all gender and age group combinations, females aged 18 to 21 years, 15 to 17 years and 22 to 25 years showed the lowest thresholds for clinically significant change, in that order. Males aged 12 to 14 years and 15 to 17 years had the highest clinically significant change indexes across all age groups and gender combinations.

### 3.2 | Expected change trajectories

Figure 1 presents the expected change trajectories by baseline MLT scores in deciles, and Table 2 shows the growth curve model slope estimates. The expected change trajectories followed a cubic pattern for those with a baseline score of 0 to 50; a quadratic pattern for baseline scores of 51 to 60; a linear pattern for baseline scores of 61 to 70; and non-significant change over time for baseline scores of 71 to 80. MLT baseline scores of 81 to 100 again followed a cubic pattern; however, this was inverse to change trajectories seen in MLT baseline scores of 0 to 50. Within baseline scores between 0 and 50, expected change trajectories for the lower scores showed a steeper increase (linear growth), greater deceleration (quadratic growth) and a bigger acceleration (cubic growth) compared with higher scores. A similar trend was evident for MLT scores between 81 and 100, but in the opposite direction, trending downwards.

### 3.3 | Early warning signals for use in clinical practice

Two early warning signals were calculated: the first was a growth curve one SD below the expected change trajectory (SD = 19.81, the yellow line in Figures 1 and 2), which would warn that the client had fallen below the 16th percentile of expected change while in treatment. The yellow line would be relevant only for MLT baseline scores of 0 to 70 as they have an increasing trend, and MLT scores for 71 to 100 would not be necessary as they would reach reliable deterioration before they dropped below one SD of the expected change trajectory. The second early warning signal (red line in Figures 1 and 2) indicates when a client has reliably deteriorated from their baseline MLT score. Reliable change has previously been calculated for MLT to be a change of 18.27 points, and reliable deterioration would mean the client has dropped 18.27 points below their baseline score (Kwan et al., 2018). The red line would be relevant for all baseline MLT scores.
Clinical MLT examples

Figures 2 and 3 provide a visual graph of the type of information that could be provided to clinicians. Currently, in headspace centres, clinicians are provided only with graphs of MLT scores over time, but the inclusion of these newly calculated benchmarks would give additional clinically useful information to help interpret the MLT scores.

The first example in Figure 2 shows a positive therapeutic change directory. It is of a 15 to 17 year old female who presented with a baseline MLT score of 24. This would mean the young person would need to reach 58.34 on MLT to obtain clinically significant change. The expected change trajectory will start at an MLT score of 24 and follow the growth curve plotted for baseline MLT scores between 21 and 30. If her score drops below the yellow line she would be one SD (SD = 19.81) below the expected change trajectory or below the

### Table 2

<table>
<thead>
<tr>
<th>Baseline MLT score</th>
<th>Slope estimates (SE)</th>
<th>Quadratic (weeks²)</th>
<th>Cubic (weeks³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10</td>
<td>5.04 (0.19)</td>
<td>0.36 (0.02)</td>
<td>0.008 (0.0006)</td>
</tr>
<tr>
<td>11-20</td>
<td>3.77 (0.11)</td>
<td>0.26 (0.01)</td>
<td>0.006 (0.0004)</td>
</tr>
<tr>
<td>21-30</td>
<td>2.82 (0.07)</td>
<td>0.17 (0.008)</td>
<td>0.004 (0.0002)</td>
</tr>
<tr>
<td>31-40</td>
<td>2.17 (0.05)</td>
<td>0.13 (0.006)</td>
<td>0.003 (0.0002)</td>
</tr>
<tr>
<td>41-50</td>
<td>1.31 (0.05)</td>
<td>0.07 (0.005)</td>
<td>0.001 (0.0002)</td>
</tr>
<tr>
<td>51-60</td>
<td>0.56 (0.03)</td>
<td>0.006 (0.0001)</td>
<td>NS</td>
</tr>
<tr>
<td>61-70</td>
<td>0.27 (0.02)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>71-80</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>81-90</td>
<td>−1.16 (0.13)</td>
<td>0.10 (0.02)</td>
<td>−0.002 (0.0005)</td>
</tr>
<tr>
<td>91-100</td>
<td>−1.83 (0.17)</td>
<td>0.14 (0.02)</td>
<td>−0.003 (0.0006)</td>
</tr>
</tbody>
</table>

Note: Slope estimates are growth curve model coefficients, Standard Error (SE), only significant estimates are shown.
16th percentile of expected change. If her MLT score further dropped equally to or below an MLT score of 5.73 during treatment (18.27 points below baseline MLT score), indicated by the red line, she would have reliably deteriorated. For this young person, her score drops to 15 in the second session, but this score is still above the yellow line, which means it is within one SD of expected change. By session seven, her MLT score is above the expected change trajectory for young people with baseline MLT scores of 21 to 30. Her progress remains above the expected change trajectory, which indicates she is making similar or better progress compared with other young people in treatment who started with a similar MLT score. At sessions 10 and 11, the young person’s MLT score is still under the clinically significant change index but her score has increased above the 18.27 points (reliable change) from her baseline indicating reliable “improvement”. By session 13, she has an MLT score of 66, which is above the clinically significant change index, meaning that this young person has moved out of the clinical population. The change can be categorized as “recovered” as the young person has reliably improved and moved from the dysfunctional population into the functional range (Jacobson & Truax, 1991).

The second example, shown in Figure 3, shows a negative therapeutic change trajectory. It is of a 12 to 14 year old female with a baseline MLT score of 36.40. The clinically significant change index would be 68.33 and her expected change trajectory would follow that of clients with baseline MLT scores between 31 and 40. By session two, this young person has a score of 20, which alerts the clinician that she has dropped below the yellow line. In the third session, the young person has a score of 26, which brings her back above the yellow line, but by session four she dips back below the yellow line with a score of 21.40. In session five, the young person has an MLT score of 10.40, which indicates she has dropped below the red line and the young person remains below the red line for the remaining sessions. In this example, the first early warning signal (yellow line) is triggered at two-time points, which tells the clinician that the client is dropping below one SD of expected change and that treatment planning may need to be reviewed. The second early warning signal (red line) is triggered by session five, showing the client has reliably deteriorated, and treatment planning and current support needed to be reviewed.

4 | DISCUSSION

The current paper aimed to develop a set of clinically significant change indexes, expected change trajectories and early warning signals to help clinicians to interpret MLT for young people aged 12 to 25 accessing youth mental health services. Using comparative scores from a nationally representative non-clinical sample, clinically significant change score benchmarks were able to be derived to assess client progress throughout treatment. Two examples were presented to demonstrate how the newly created benchmarks and early warning signals could be used to inform clinical practice. Table 3 summarizes the clinical benchmarks for MLT, which in conjunction with the other
These indexes are calculated by finding the value where the non-clinical and clinical populations intersect. MyLifeTracker has clinically significant change indexes based on gender and age group (see Table 1). When a reliable change (18.27 points) is also considered, change can be categorized into four stages:
- Recovered—when a client reliably improves and moves into the functional population
- Improvement—when a client has made a reliable positive change but still remains in the dysfunctional population
- Unchanged—when no reliable change has occurred
- Deterioration—when a client has reliably worsened (see below in “Early warning signals—Red line” section)

Note: If a client is above the clinically significant change index, a client cannot reach “recovered” and it may be difficult to achieve reliable “improvement” due to how high the client’s score is and because they are already more likely to be in the functional population. The client can still show reliable “deterioration”.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinically significant change index</td>
<td>This index provides clinicians with information on whether a client is more likely to be in the non-clinical (above the index) or clinical population (below the index). It allows clinicians to see when a client moves from the dysfunctional to the functional population during treatment, known as “clinically significant change”. These indexes are calculated by finding the value where the non-clinical and clinical populations intersect. MyLifeTracker has clinically significant change indexes based on gender and age group (see Table 1). When a reliable change (18.27 points) is also considered, change can be categorized into four stages: Recovered—when a client reliably improves and moves into the functional population Improvement—when a client has made a reliable positive change but still remains in the dysfunctional population Unchanged—when no reliable change has occurred Deterioration—when a client has reliably worsened (see below in “Early warning signals—Red line” section) Note: If a client is above the clinically significant change index, a client cannot reach “recovered” and it may be difficult to achieve reliable “improvement” due to how high the client’s score is and because they are already more likely to be in the functional population. The client can still show reliable “deterioration”.</td>
</tr>
<tr>
<td>Expected change trajectory</td>
<td>This trajectory provides clinicians with estimates of average rates of change for clients. An individual client’s trajectory can be compared with the average trajectory to determine whether the client is within or outside expected rates of change. These trajectories are calculated using growth curve modelling based on a clinical group during an episode of care. MyLifeTracker has expected change trajectories calculated for decile groups dependent on MyLifeTracker baseline scores, that is, 0-10, 11-20, etc (see Table 2 and Figure 1).</td>
</tr>
<tr>
<td>Early warning signals—Yellow line</td>
<td>This yellow line provides clinicians with a warning when a client drops one SD below the expected change trajectory. This would mean that the client has fallen below the 16th percentile of expected change while in treatment and that treatment planning may need to be reviewed. These yellow lines are modelled on the same growth curve as the expected change trajectories, however, start from 19.81 points (one SD) below the client’s MyLifeTracker baseline score (see examples in Figures 2 and 3). Note: The yellow line for MyLifeTracker is only relevant for baseline scores of 0-70 as they have an increasing trend, and scores for 71-100 are not necessary as they would reach reliable deterioration (red line) before they dropped below one SD of the expected change trajectory.</td>
</tr>
<tr>
<td>Early warning signals—Red line</td>
<td>This red line provides clinicians with a warning when a client has reliably deteriorated from their baseline score during their course of treatment. This may indicate that the client has increased risk or concerns, is not responding to treatment and may prematurely dropout from treatment. Clinicians should review treatment planning and check if additional supports are required. These red lines are calculated as 18.27 points (reliable change) below the client’s MyLifeTracker baseline score (see examples in Figures 2 and 3). Note: The red line would be relevant for all MyLifeTracker baseline scores. However, the red line will not exist when the MyLifeTracker baseline score is too low as a client’s score cannot drop below 0 during treatment.</td>
</tr>
</tbody>
</table>

Tables and figures, provides a quick guide of how clinicians can use and interpret MLT.

As hypothesized, clinically significant change indexes were distinctly different across age groups and gender, with male adolescents showing a higher threshold by seven to 10 MLT points compared to female adolescents. While young adult males also showed a higher threshold than young adult females, this gap was smaller (two to four MLT points). Overall, the largest difference between indexes was between the males aged 12 and 14 years and females aged 18 and 21 years, with a difference of 18 MLT points. This clearly demonstrates the need for gender and age-specific clinically significant change indexes to provide appropriate benchmarks responsive to the distinct developmental variances occurring during this rapidly changing time of life (Donald, Carey, & Rickwood, 2018; McGorry, Goldstone, Parker, Rickwood, & Hickie, 2014).

The expected change trajectories followed a cubic pattern for MLT baseline scores below 50, and this pattern of change has been demonstrated in other naturalistic settings (Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009). It was shown that patterns of change were faster with lower MLT baseline scores, compared to higher MLT baseline scores below 50, and this pattern has been shown with other studies exploring youth outcomes (Cannon et al., 2010). These models of change using MLT are consistent with previous research that suggests there is a likelihood of more sudden change early in treatment and then a deceleration as treatment progresses (Baldwin et al., 2009; Gaynor et al., 2003; Tang, Luborsky, & Andrusyna, 2002). There is an increasing, but slower, rate of change among baseline MLT scores between 51 and 70, whereas baseline MLT scores of 71 to 80 showed no change over time. This can be explained by MLT scores being closer to the clinically significant change indexes, and it is expected that there will be less change over time as clients are already closer to the functional distribution. Baseline MLT scores of 81 to 100 showed an inverse cubic pattern to those baseline MLT scores under 50, with the MLT scores declining, and this ceiling effect is common across outcome measures for clients that rate their mental health very positively (Higginson & Carr, 2001).

The results of this study add to the growing research towards increasing the utility of youth mental health outcome measures to...
Young people are shown to have higher rates of deterioration and clinicians are shown to have lower rates of accurately predicting deterioration compared to adults in mental health treatment (Cannon et al., 2010; Warren et al., 2009). They are also more likely to show higher treatment dropout and missed appointments, and it has been suggested that this is due to their perceptions around the usefulness of professional help and stigma related to this (O’Brien, Fahmy, & Singh, 2009). This higher level of disengagement is particularly seen with young people who are males, Aboriginal or Torres Strait Islander, aged over 18 years and living in rural areas. However, a high number of those who discontinue from treatment are shown to reengage in the future, and those young people may need to engage multiple times to meet their mental health needs (Seidler et al.). As such, the use of MLT in FIT targets the developmental period spanning the 12 to 25 age range that may be quite responsive to this type of monitoring during treatment (Donald et al., 2018; Langer & Jensen-Doss, 2018).

The clinically significant change indexes, expected change trajectories and early warning signals developed here provide important information to help youth mental health clinicians interpret changes in MLT scores. The functionality to include these indexes in the current headspace data collection system is not yet available, although sophisticated electronic measurement systems, tailored to clients’ age and gender and baseline outcome scores, are becoming available. Such information can be very helpful for clinicians, to inform clinical practice and provide feedback to clients, and also clinicians’ own deliberate practice. Deliberate practice, which is a process of systematic effort to improve performance with the guidance of a supervisor, ongoing feedback relative to essential skills, and refinement and repetition of practice (Goodyear et al., 2017), has been shown to contribute to differences between clinicians in client outcomes, with the most effective clinicians engaging in 2.8 times more deliberate practice than other clinicians (Chow et al., 2015).

There are still mixed views among clinicians using FIT, however, and this seems to affect its effectiveness (Lucock et al., 2015; Lutz et al., 2015). De Jong, Van Sluis, Nugter, Heiser, and Spinhoven (2012) showed clinicians who used the measurement feedback provided to them had improved outcomes for those clients NOT. Specifically, female clinicians and clinicians reporting higher commitment to using FIT at the start of treatment were more likely to use the feedback provided from the measure. Further, clinicians who were more likely to trust feedback from sources external to their own opinion (low internal feedback propensity), had clients with faster rates of change compared to clinicians with a high internal feedback propensity. Clinicians with a strong focus on achieving success (promotion focussed) were more likely to achieve better outcomes using feedback when compared to clinicians who focus on preventing failures (prevention focussed) (De Jong & De Goede, 2015). At a service level, clinics that showed a better implementation of feedback systems were more likely to have measures completed and outcomes viewed by clinicians, which in turn led to a more positive impact on client outcomes (Bickman et al., 2016). Training is increasingly available in the area of FIT and future research should target how to improve clinicians’ acceptability of feedback monitoring systems and how to enhance its implementation and effectiveness (Law & Wolpert, 2014).

The results of the current study should be interpreted in light of its limitations. Notably, the clinically significant change indexes, expected change trajectories and early warning signals were created for an early intervention mental health service for young people aged 12 to 25 years in Australia. Further research is needed to determine whether the benchmarks would apply to young people attending specialist or tertiary services. The indexes were developed using a community sample from Australia, and it is unknown whether similar MLT scores would be found in other countries. Replication in other regions of the world focusing on the development of youth mental health systems, like Canada, Ireland, the Netherlands and California, is warranted (McGorry, Trethowan, & Rickwood, 2019). Furthermore, the current study only explored expected change trajectories dependent on baseline MLT scores, as past studies have shown that this accounts for a significant amount of variance in the rate of change (Lambert et al., 2002). However, it may be important also to create expected change trajectories for other predictors, such as the client’s diagnosis and presenting issues. For example, a study on substance abuse treatment found that while baseline mental health measures were a significant predictor of rates of change, employment status and baseline craving levels were also significant predictors of rates of change (Crits-Christoph et al., 2015).

In conclusion, the development of these MLT benchmarks is an important step to increase the clinical utility of the measure. MLT was originally developed to fill a gap in the availability of routine outcome measures for youth mental health services provided to adolescents and young adults. The availability of these benchmarks, including clinically significant change indexes and expected change trajectories, enhances the clinical utility and interpretability of the measure (Boswell et al., 2015; Donald & Carey, 2017). Providing benchmarks that are age group and gender specific is also critical for this age range when there is a substantial developmental change occurring in multiple domains. The clinical benefits of FIT are becoming more widely known and have become part of the agenda for the future progression of psychotherapy (Emmelkamp et al., 2014; Lutz et al., 2015). It is essential that such practices can be applied in youth mental health, where dropout and lack of clinical change are particularly problematic. The implementation of routine outcome measures, like MLT, and the use of benchmarks that enable clinicians to determine developmentally appropriate change directories that reveal recovery, improvement, lack of change or deterioration, is essential to supplement clinical judgement to improve clinical practice and outcomes in youth mental health settings.

ACKNOWLEDGEMENTS

The authors acknowledge the work of the headspace programme and services, and headspace staff and clients in providing data. headspace National Youth Mental Health Foundation is funded by the Australian Government.
CONFLICT OF INTEREST
All authors are either employed by headspace National or a headspace centre. The authors report no other conflicts of interest in this work.

DATA AVAILABILITY STATEMENT
The data are available only to the collaborating researchers with the headspace National Youth Mental Health Foundation. Some data may be available upon reasonable request, but not all due to relevant data protection laws.

ORCID
Benjamin Kwan https://orcid.org/0000-0001-9532-2990
Debra J. Rickwood https://orcid.org/0000-0002-4227-0231

REFERENCES


---

How to cite this article: Kwan B, Rickwood DJ. A routine outcome measure for youth mental health: Clinically interpreting MyLifeTracker. Early Intervention in Psychiatry. 2020;1–11. https://doi.org/10.1111/eip.13016
Chapter VI: Factors Affecting the Implementation of an Outcome Measurement Feedback System in Youth Mental Health Settings

Chapter VI presents the fourth and final paper prepared during the PhD candidature. The paper aimed to explore factors affecting the use of MLT within MFS across youth mental health settings in Australia. Paper Four has been peer reviewed and was published in *Psychotherapy Research* in October 2020. *Psychotherapy Research* is a Q1 journal in the Clinical Psychology category with an h-index of 71 and an impact factor of 2.984. The format of the paper is in accordance with the publication guidelines for this journal.
# Declaration of Co-Authored Publications

For use in theses which include co-authored publications. This declaration must be completed for each co-authored publication and to be placed at the start of the thesis chapter in which the publication appears, or as a preface to the thesis.

## Declaration for Thesis Chapter VI

**DECLARATION BY CANDIDATE**

In the case of Chapter VI, the nature and extent of my contribution to the work was the following:

<table>
<thead>
<tr>
<th>Nature of Contribution</th>
<th>Extent of Contributions (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptualisation and design of the study, review of the literature, data analysis, and write up of the article.</td>
<td>75%</td>
</tr>
</tbody>
</table>

The following co-authors contributed to the work:

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
<th>Contributor is also a UC student (Yes/No)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debra Rickwood</td>
<td>Conceptualisation and design of the study, review of the results, and editing the article.</td>
<td>No</td>
</tr>
<tr>
<td>Patricia Brown</td>
<td>Conceptualisation and design of the study, review of the results, and editing the article.</td>
<td>No</td>
</tr>
</tbody>
</table>

Candidate’s Signature: [Signature]

Date: 18/10/2020

**DECLARATION BY CO-AUTHORS**

The undersigned hereby certify that:

1. the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. there are no other authors of the publication according to these criteria;
5. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

<table>
<thead>
<tr>
<th>Location(s):</th>
<th>University of Canberra</th>
</tr>
</thead>
</table>

Signatures

<table>
<thead>
<tr>
<th>[Signature]</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>19/10/2020</td>
</tr>
<tr>
<td></td>
<td>22/10/2020</td>
</tr>
</tbody>
</table>
EMPIRICAL PAPER

Factors affecting the implementation of an outcome measurement feedback system in youth mental health settings

BENJAMIN KWAN 1, DEBRA J. RICKWOOD 12, & PATRICIA M. BROWN 1

1Faculty of Health, University of Canberra, Bruce, ACT, Australia & 2Headspace National Youth Mental Health Foundation National Office, Melbourne, VIC, Australia

(Received 27 April 2020; revised 18 September 2020; accepted 21 September 2020)

Abstract
Objective: Measurement feedback systems provide clinicians with regular snapshots of a client’s mental health status, which can be used in treatment planning and client feedback. There are numerous barriers to clinicians using outcome measures routinely. This study aimed to investigate factors affecting the use of a measurement feedback system across youth mental health settings. Methods: The participants were 210 clinicians from Headspace youth mental health services across Australia. They were surveyed on predictors and use of MyLifeTracker, a routine outcome measure. This was explored through three processes: looking at MyLifeTracker before session, using MyLifeTracker in treatment planning, and providing feedback of MyLifeTracker scores to clients. Results: Clinicians were more likely to look at MyLifeTracker before session, less likely to use it in treatment planning, and least likely to provide MyLifeTracker scores to clients. Each measurement feedback system process had a distinct group of predictors. Perceptions of MyLifeTracker’s practicality was the only significant predictor of all three processes. Conclusion: Practically, organisations and supervisors can increase the use of measurement feedback systems through targeted supports.

Keywords: MyLifeTracker; measurement feedback systems; youth mental health; routine outcome measure; progress monitoring

Clinical or Methodological Significance: The study investigates the factors that affect the implementation of measurement feedback systems in a naturalistic youth mental health setting. It examines the use of MyLifeTracker, a session-by-session mental health outcome measure, across Headspace centres in Australia. The study identifies clinician characteristics, clinician attitudes, and organisational supports as important predictors for clinicians use of measurement feedback systems. Specifically, it provides an in-depth examination of three distinct measurement feedback system processes: looking at MyLifeTracker before session, using MyLifeTracker in treatment planning, and providing feedback of MyLifeTracker scores to clients. Young people have high rates of deterioration during therapy, treatment drop-out, and missed appointments. As such, the implementation of MyLifeTracker in measurement feedback systems targets a critical developmental period where this type of monitoring and feedback during treatment may be particularly useful.

The development of session-by-session client outcome measures has enabled the implementation of measurement feedback systems (MFS), also known as routine outcome monitoring (ROM), progress monitoring, and feedback-informed treatment (FIT). These are platforms that allow clinicians to receive quantified feedback on a client’s progress to use in-session and help guide treatment planning (Boswell et al., 2015; Miller et al., 2015). MFS require a process that is efficient for the client and allows instant feedback to the clinician. Clinicians receive regular up-to-date snapshots of a client’s
mental health status, showing any changes since past sessions. Clinicians can monitor if clients are progressing or deteriorating between sessions and adjust treatment planning accordingly (Boswell et al., 2015). Such an approach can also allow clinicians to discuss measurement results in session and provide progress feedback to clients, which can be a powerful way to promote shared decision making (De Jong et al., 2014). MFS have been shown to improve communication between client and clinician, increase the accuracy of diagnosis, enable quicker adjustments to treatment planning when required, provide stronger outcome effects, and improve the efficiency of treatment (Bickman et al., 2011; Carlier et al., 2012; Janse et al., 2017). MFS are particularly useful for clients “not on track” (NOT) or who are at risk of treatment failure, with feedback systems significantly reducing deterioration rates and increasing rates of clinically significant improvement (Shimokawa et al., 2010).

MFS are important because clinicians have been shown to have low accuracy rates in predicting client deterioration during therapy when using their judgement alone (Hannan et al., 2005; Hatfield et al., 2010). They tend to have a self-assessment bias which serves to maintain a positive self-image by overestimating treatment progress (Parker & Waller, 2015). Walfish et al. (2012) explored clinicians’ ratings of their own clinical skills and client outcomes, showing that they rated their skills on average at the 80th percentile and that all clinicians rated themselves above the 50th percentile. Additionally, clinicians on average believed that 77% of their clients improved as a result of their therapeutic intervention, which is well above the one-third proportion of clients shown to improve in naturalistic settings (Walfish et al., 2012). Formal quantified feedback provides a structured method for reducing this self-assessment bias (Macdonald & Mellor-Clark, 2015), and deliberate practice, incorporating the use of MFS, can improve the effectiveness of clinicians (Chow et al., 2015).

Despite the benefits of MFS becoming more widely known and increased training in the use of MFS in clinical programmes (Overington et al., 2015), clinicians still have a typically low completion rate of routine outcome measures, which suggests difficulties with the administration of measures and their clinical use (Batty et al., 2013; Hatfield & Ogles, 2004; Johnston & Gowers, 2005). Interestingly, clinicians that primarily used Cognitive Behavioural Therapy (CBT) were more likely to view outcome measures as clinically useful and practical compared to practitioners of other therapeutic orientations (Jensen-Doss et al., 2018). Clinicians who had recently graduated were also more likely to be aware of routine outcome measures compared to those with more years of experience, while clinicians with higher-level degrees were more likely to use routine outcome measures compared to those with lower-level degrees (Ionita & Fitzpatrick, 2014). Clinicians in private practice were less likely to administer outcome measures, saw less benefit in monitoring and feedback, and were less likely to find measures clinically useful and practical compared to clinicians working in institutional settings (Hatfield & Ogles, 2004; Jensen-Doss et al., 2018).

One of the strongest predictors of any outcome measure use was fewer years of professional experience, which was a significant unique predictor even when other demographic characteristics, professional characteristics, and attitudes towards MFS were accounted for (Jensen-Doss et al., 2018).

Some clinicians are not convinced of the effectiveness of MFS and believe that the measures are not as accurate as general clinical judgement (Ross et al., 2016). Importantly, clinicians have raised concerns around the clinical usefulness and practicality of measures used in MFS. They report that some measures can take too long to administer, lack timely feedback, or it is challenging to access results (Ionita et al., 2016; Ross et al., 2016). Other concerns are lack of relevance of measurement items for clinicians and clients, and lack of sensitivity to change (Bickman, 2008; Happell, 2008; Kelley & Bickman, 2009). There is a strong association between clinicians’ attitudes toward measure usefulness and MFS use (Chung & Buchanan, 2019). Jensen-Doss and colleagues (2018) explored clinicians’ attitudes to standardised measures and found that clinicians who were more positive about a measure’s clinical utility, treatment planning properties, and practicality were more frequently using routine outcome measures. The practicality of the measure was shown to be the strongest predictor of frequent use.

Service and organisation factors also play a vital role in the successful implementation of routine outcome measures and MFS. When a workplace mandates the use of progress monitoring, clinicians hold more positive views towards outcome measures and are more likely to administrator them (Jensen-Doss et al., 2018). More specifically, organisational resources that incorporate technology to reduce administrative burden and increase the timeliness of feedback can help increase the use of MFS (Bickman et al., 2016). Another organisational support is the provision of appropriate training in MFS, which can increase clinicians’ positive attitudes towards MFS and self-efficacy in use (Edbrooke-Childs et al., 2016). A trial of MFS in two youth clinics revealed that the clinic that provided more organisational supports engaged in MFS more often and had better treatment outcomes (Bickman et al.,
While both clinics had the same training protocols and barriers to using new technology systems, organisational supports targeting championing of MFS and day-to-day clinical supervision were seen to be the critical factors for successful implementation of MFS.

Ross and colleagues (2016) studied the implementation of a secure web-based platform named Client-Reported Outcomes Monitoring Information Systems (CROMIS) rolled out across Operational Stress Injury clinics in Canada. The system included several measures that clinicians could choose as appropriate for clients. Clients could complete them electronically, and results were automatically scored with change graphs instantly available to clinicians. The rollout included training, technical support, and opportunities for comments. The study found that 85% of clinicians administering measures, 78% reviewing results, and 63% discussing results with clients. There were no differences in barriers between users and non-users when administering measures; however, non-users perceived “burden” and “organisational supports” as more prominent barriers when reviewing measure results. “Burden” was also a more significant barrier for non-users when discussing results with clients. These results suggest that there may be different barriers and facilitators to the different processes of MFS.

MyLifeTracker in headspace Centres

MyLifeTracker (MLT) is a routine outcome measure designed for youth mental health settings (Kwan et al., 2018). It is brief, with only five items, making it suitable for session-by-session use, and was co-designed with young people and youth mental health clinicians to assess meaningful outcomes. It assesses the self-reported current quality of life in the domains of general wellbeing, day-to-day activities, relationships with friends, relationships with family, and coping. It is psychometrically sound across the 12–25 year age range, displaying appropriate validity, reliability, and sensitivity to change. It was developed for this age range to respond to the growing international movement toward youth mental health services, necessitating the creation of new measures suitable for both adolescents and young adults (Kwan & Rickwood, 2015; McGorry et al., 2013). It has a single factor structure, although the five items were also designed to be clinically useful individually (Kwan et al., 2018).

MLT is currently implemented via an electronic data system within more than 110 headspace youth mental health centres throughout Australia. headspace is the Australian Government’s National Youth Mental Health Foundation, initiated in 2006 to provide early intervention in youth mental health for those aged 12–25 years (Rickwood et al., 2018). headspace centres offer holistic primary care services responding to mental health, alcohol and other drugs, physical/sexual health, and vocational concerns for young people. MLT is administered through a tablet device or computer before a young person attends each clinical session. The measure is sent to the young person via email or text message or completed while they are waiting in reception. New clients to headspace are provided with an online consent form when they first log in to the headspace data collection system. This advises why headspace collects the information and how it is used (Rickwood et al., 2018). The self-reported young person’s results are instantly available to the clinician prior to the session in the form of a graph over time. The graph shows every session the young person has attended and can be viewed as a total MLT score or as individual items.

Young people have higher rates of deterioration during therapy and clinicians have lower rates of accurately predicting deterioration with young people, compared to adults in mental health treatment (Cannon et al., 2010; Warren et al., 2009). They are also more likely to show higher rates of treatment drop-out and missed appointments, and it has been suggested that this is due to their perceptions around the usefulness of professional help and stigma related to this (O’Brien et al., 2009). This higher level of disengagement is more common with young people who are males, Aboriginal or Torres Strait Islander, aged over 18 years, and living in rural areas. However, a high number of those who discontinue from treatment are shown to reengage in the future, and young people may need to engage multiple times to meet their mental health needs (Seidler et al., 2020). As such, the use of MLT in MFS targets a critical developmental period where this type of monitoring during treatment may be particularly useful (Donald et al., 2018; Langer & Jensen-Doss, 2018).

The Current Study

The current study investigates the factors that affect the implementation of MFS in a naturalistic youth mental health setting by examining the usage of MLT in headspace centres. The first aim was to describe how much clinicians are using MLT in MFS. It was hypothesised that clinicians would be most likely to look at MLT before session, less likely to use MLT in treatment planning, and least likely to provide feedback of MLT scores to clients.
The second aim was to examine the nature of potential predictors of MLT use in MFS, specifically: attitudes to MFS in general; attitudes to MLT specifically; and perceptions of organisational support for MLT. It was hypothesised that clinicians would hold neutral to positive attitudes and perceptions towards these factors based on previous research exploring the use of MFS (Jensen-Doss et al., 2018).

The third aim was to identify the strongest predictors of MLT use in MFS, specifically: clinician characteristics; attitudes to MFS in general; attitudes to MLT specifically; and perceptions of organisational support for MLT. It was hypothesised that fewer years of clinical experience, positive attitudes towards MLT practicality, positive perceptions of organisation championing for MLT, and positive perceptions of organisational clinical support for MLT would be the strongest predictors of MLT use in MFS.

Method

Participants

Participants comprised 210 clinicians from 88 headspace centres across Australia. Participant demographic and professional characteristics are presented in Table I. Participants were mostly female, psychologists, had a bachelor’s or master’s degree, used cognitive behavioural therapies, and were employed as salaried staff.

Procedure

Participants were recruited through an email invitation sent to 1176 clinicians in headspace centres across Australia. Those interested in participating were directed to a link connected to the online Qualtrics platform that provided participant information and then the survey questionnaire. The response rate was 17.9%. Data were collected between 10th April 2019 and 31st July 2019. Ethics approval was obtained from the University of Canberra Human Research Ethics Committee (Project no. 0468).

Measures

The Qualtrics survey collected demographic information as reported in Table I and the following measures.

**MyLifeTracker measurement feedback system use (MLT use in MFS).** To assess clinician use of MLT in MFS, three items were adapted from previous research to gauge the use of MLT in each of the different processes within MFS (Lyon et al., 2019; Ross et al., 2016). Participants were asked to indicate the percentage of sessions in which they use MLT in the following ways:

- look at MyLifeTracker before session?
- use MyLifeTracker to inform treatment planning in any way (i.e., alter plan, continue plan as it supports it, add more supports)?
- provide feedback of MyLifeTracker scores to clients in any way (i.e., discuss results, show graph results, ask about areas of change/deterioration or no change)?

**Monitoring and feedback attitudes scale (MFA).** To assess clinician attitudes towards MFS, including the processes of administration of routine measures, use of measures in treatment planning, and feedback of treatment progress to clients, the
14-item Monitoring and Feedback Attitudes Scale (MFA) was used (Jensen-Doss et al., 2018). In the MFA instructions, definitions of routine progress monitoring and providing feedback were included. Participants were asked to indicate how much they agreed or disagreed with each statement on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree). The MFA consists of two subscales: MFA Benefit (10 items) and MFA Harm (4 items). Total subscale scores were calculated by averaging across the items with higher subscale scores indicating higher levels of endorsement for monitoring and feedback benefits and harm. Jensen-Doss et al. (2018) found this measure to have good internal consistency (MFA Benefit $\alpha = 0.87$, MFA Harm $\alpha = 0.87$). The current study also found good internal consistency (MFA Benefit $\alpha = 0.93$, MFA Harm $\alpha = 0.82$).

Attitudes toward MyLifeTracker—monitoring and feedback (AM-MF). To assess clinician attitudes toward using MLT for clinical decision making and feedback, 18 items were adapted from the Attitudes Toward Standardized Assessment Scales—Monitoring and Feedback (ASA-MF; Jensen-Doss et al., 2018). The phrase “standardised progress measures” was replaced with “MyLifeTracker” to focus on views of MLT specifically. The wording “administering” measures was replaced by “accessing and reviewing” measures for relevant items, as the administration of MLT is completed by reception staff in this study. Participants were provided with the definition of routine progress monitoring, directed to answer the items based specifically on MLT, and provided visual examples of the MLT measure and graphs. They indicated how much they agreed or disagreed with each statement on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree). Total subscale scores were calculated by averaging across the items, with higher subscale scores indicating more positive attitudes towards MLT. The AM-MF consisted of three subscales, all with acceptable internal consistency: Clinical Utility (8 items, $\alpha = 0.79$), Treatment Planning (5 items, $\alpha = 0.79$), and Practicality (5 items, $\alpha = 0.77$). Jensen-Doss et al. (2018) found the original ASA-MF to have good internal consistency (Clinical Utility $\alpha = 0.85$, Treatment Planning $\alpha = 0.85$, and Practicality $\alpha = 0.81$).

Organisational factors for MyLifeTracker use scale (OFM). To assess clinician perceptions of organisational factors in the use of MLT, 21 items were adapted from previous research exploring organisational factors for routine outcome measure use (Gleacher et al., 2016; Jensen-Doss et al., 2018; Ross et al., 2016). Items covered organisation resources such as, “I have adequate technological resources to use MyLifeTracker for routine progress monitoring and providing feedback (e.g., computer or tablet devices or printers)”; championing of MLT, “the organisation encourages me to use MyLifeTracker with treatment planning and providing feedback within sessions”; clinical support, “my supervisor/manager provides clinical support of MyLifeTracker with treatment planning and providing feedback within sessions”; and training, “I have adequate training with using MyLifeTracker in treatment planning.” Definitions of routine progress monitoring and providing feedback were again included, and participants were again directed to answer the items based specifically on MLT. They indicated how much they agreed or disagreed with each statement on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree). Total subscale scores were calculated by averaging across the items with higher subscale scores indicating more positive perceptions of organisational factors in the use of MLT. The OFM consisted of four subscales, all with good internal consistency: Resources (6 items, $\alpha = 0.86$), Championing (7 items, $\alpha = 0.85$), Clinical Support (3 items, $\alpha = 0.83$), and Training (5 items, $\alpha = 0.90$).

Data Analyses

Levels of use of each of the three MFS processes were the dependent variables. These were the percentage of client sessions in which participants used MLT in the following ways: looked at MLT before session, used MLT in treatment planning, and provided feedback of MLT scores to clients. A one-way repeated measures analysis of variance (ANOVA) was used to compare the percentage of client sessions participants used MLT across the three MFS processes. The proportion of participants who marked 0% use to each MFS process was also calculated. A Cochran’s $Q$ test was used to test for differences between related proportions of participants who marked 0% use across the three MFS processes.

For analysis purposes, the following clinician demographics were included as predictor variables and were transformed into binary variables: highest education level — Diploma/Bachelor’s (0), Masters/Doctoral (1); primary therapeutic orientation — All Else (0), Cognitive Behavioural Therapy (1); funding source — Private Practitioner/In-Kind Partner/Other (0), On-Staff Grant Funded (1). Years of clinical experience was also included in analyses. The subscales of MFA, AM-MF, and OFM were the remaining predictor variables.

To examine whether the potential predictors of MLT use in MFS were positively or negatively valenced, a one-sample $t$-test was used to compare
the subscale scores of MFA, AM-MF, and OFM against a neutral rating of 3. Cohen’s $d$ effect sizes were computed to identify the magnitude of attitude and perception strength. Correlational analysis examined relationships among predictors and dependent variables. Multiple linear regression was used to assess the relative contributions of the predictor variables to explaining the variance of the three dependent variables. Effect sizes were interpreted according to Cohen’s conventions (Cohen, 1988). SPSS V25 was used for all quantitative analyses.

### Results

**Clinician MLT Use in MFS**

Overall, participants looked at MLT before just over half ($M = 52.48\%$, $Median = 55.50\%$, $SD = 37.87\%$) of their sessions, used MLT in treatment planning for slightly more than one-third ($M = 36.23\%$, $Median = 32.50\%$, $SD = 29.60\%$), and provided feedback of MLT scores to clients in a quarter ($M = 24.67\%$, $Median = 16.00\%$, $SD = 25.86\%$) of sessions. A repeated-measures ANOVA with a Greenhouse-Geisser correction determined that the average percentage of client sessions participants used MLT across the three MFS processes differed significantly with a large effect size, $F (1.74, 280.01) = 75.62$, $p < .001$, $\eta^2_p = .32$. Pairwise comparisons using the Bonferroni correction revealed that each MFS process differed significantly from the others ($p's < .001$).

The proportion of participants that reported that they did not use MLT in MFS, specifically marking 0% of MLT use in sessions were 13.3% (before session), 15.2% (treatment planning), and 24.1% (client feedback). Cochran’s $Q$ (with $\alpha = .05$), which tests for differences in related proportions, was statistically significant, $Q (2, N = 162) = 18.24$, $p < .001$. To locate the source of this significance, a series of pairwise comparisons using the McNemar test with the Bonferroni correction ($\alpha = .017$) were undertaken. The proportion of participants that did not use client feedback was significantly higher than those who did not look at MLT before session ($p < .001$) and did not use it in treatment planning ($p = .003$), respectively. The proportion of participants that did not look at MLT before session and did not use it in treatment planning were not significantly different ($p = .36$).

**Clinician Attitudes and Perceptions for MFS and MLT**

Table II contains the subscale scores for MFA, AM-MF, and OFM, and their effect sizes when compared to a neutral rating of 3. The MFA Benefit subscale was on average positive and showed a large effect size when compared to a neutral rating, while participants disagreed with the MFA Harm subscale with a moderate to large effect size. Participants were generally positive about MLT with a small to medium effect size for AM-MF Clinical Utility, a medium effect size for AM-MF Treatment Planning, and a large effect for AM-MF Practicality. Participants on average were positive about OFM Resources with a large effect size, while they were negative about OFM Clinical Support with a moderate effect size. Differences were not significant for OFM Championing and OFM Training.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>$M (SD)$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFA Benefit</td>
<td>4.17 (0.49)*</td>
<td>2.39</td>
</tr>
<tr>
<td>MFA Harm</td>
<td>2.48 (0.66)*</td>
<td>−0.79</td>
</tr>
<tr>
<td>AM-MF Clinical Utility</td>
<td>3.20 (0.53)*</td>
<td>0.38</td>
</tr>
<tr>
<td>AM-MF Treatment Planning</td>
<td>3.31 (0.62)*</td>
<td>0.50</td>
</tr>
<tr>
<td>AM-MF Practicality</td>
<td>3.57 (0.63)*</td>
<td>0.90</td>
</tr>
<tr>
<td>OFM Resources</td>
<td>3.75 (0.71)*</td>
<td>1.06</td>
</tr>
<tr>
<td>OFM Championing</td>
<td>2.98 (0.67)*</td>
<td>−0.03</td>
</tr>
<tr>
<td>OFM Clinical Support</td>
<td>2.40 (0.86)*</td>
<td>−0.70</td>
</tr>
<tr>
<td>OFM Training</td>
<td>2.90 (0.94)*</td>
<td>−1.11</td>
</tr>
</tbody>
</table>

Table II. Subscale scores for MFA, AM-MF, and OFM, and their effect sizes when compared to a neutral rating of 3.

**Intercorrelations Among Predictors and MLT Use in MFS**

Table III presents the intercorrelations between predictor and dependent variables. Fewer years of clinical experience was related to higher levels of looking at MLT before session. Education level and funding source were not related to any MLT use in MFS. CBT-oriented clinicians were slightly more likely to look at MLT before session and provide feedback on MLT scores to clients. The MFA benefit subscale was weakly associated with the use of MLT in treatment planning and providing feedback on MLT scores to clients. Participants who disagreed with items on the MFA harm subscale were weakly more likely to use MLT in treatment planning. All three AM-MF subscales and all four OFM subscales were positively correlated to all three processes of MLT use in MFS, with AM-MF Practicality and OFM Championing showing the strongest relationships. All three dependent variable scales for MLT use in MFS were moderate to strongly intercorrelated.

**Predictors of MLT use in MFS**

Table IV presents the multiple regression models for predicting use of MLT in MFS: specifically looking
Predictors of looking at MLT before session.

The model explained 41.6% of the variance in clinicians looking at MLT before session, adjusted $R^2 = .36$, $F(13, 147) = 8.05$, $p < .001$, a large effect ($f^2 = 0.71$). Significant predictors were years of clinical experience, AM-MF clinical utility subscale, and OFM championing, uniquely explaining 2.3% and 1.0% of the variance, respectively. The MFA benefit subscale was not statistically significant, explaining 3.2% of the variance.

Predictors for using MLT to inform treatment planning.

The model explained 45.8% of the variance in clinicians using MLT to inform treatment planning, adjusted $R^2 = .41$, $F(13, 147) = 9.43$, $p < .001$, a large effect ($f^2 = 0.85$). Significant predictors were AM-MF treatment planning, AM-MF practicality, and OFM championing, uniquely explaining 2.3% and 0.8% of the variance, respectively.

**Table III. Intercorrelations among predictor and dependent variables.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years Clin Exp</td>
<td>-.15*</td>
<td>-.21**</td>
<td>-.20**</td>
<td>.00</td>
<td>-.21**</td>
<td>.03</td>
<td>.01</td>
<td>-.08</td>
<td>-.17*</td>
<td>-.04</td>
<td>-.04</td>
<td>-.03</td>
<td>-.26**</td>
<td>-.11</td>
<td>-.04</td>
<td></td>
</tr>
<tr>
<td>Ed Level</td>
<td>-.15*</td>
<td>.01</td>
<td>-.06</td>
<td>-.04</td>
<td>.03</td>
<td>-.08</td>
<td>-.02</td>
<td>.01</td>
<td>-.05</td>
<td>-.04</td>
<td>-.03</td>
<td>-.06</td>
<td>-.11</td>
<td>-.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funding Source</td>
<td>-.01</td>
<td>.06</td>
<td>.05</td>
<td>-.07</td>
<td>.04</td>
<td>.04</td>
<td>.09</td>
<td>.05</td>
<td>.06</td>
<td>.01</td>
<td>.09</td>
<td>.08</td>
<td>-.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Therapy Orient</td>
<td>.01</td>
<td>-.09</td>
<td>-.09</td>
<td>.04</td>
<td>-.08</td>
<td>.10</td>
<td>.15*</td>
<td>.04</td>
<td>-.02</td>
<td>-.02</td>
<td>.00</td>
<td>.18*</td>
<td>.10</td>
<td>.18*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFA Benefit</td>
<td>-.24**</td>
<td>.24**</td>
<td>-.37**</td>
<td>.31**</td>
<td>.14</td>
<td>.56**</td>
<td>.19*</td>
<td>.15*</td>
<td>.08</td>
<td>.21**</td>
<td>.20*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM-MF Ham</td>
<td>-.31***</td>
<td>-.59**</td>
<td>.55**</td>
<td>.14</td>
<td>.33**</td>
<td>.13</td>
<td>.09</td>
<td>.16</td>
<td>.38**</td>
<td>.16*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM-MF CU</td>
<td>-.48**</td>
<td>-.48**</td>
<td>.42**</td>
<td>.47**</td>
<td>.19</td>
<td>.17</td>
<td>.26**</td>
<td>.49**</td>
<td>.24**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM-MF TP</td>
<td>-</td>
<td>.41***</td>
<td>.39**</td>
<td>.13</td>
<td>.27**</td>
<td>.51**</td>
<td>.50**</td>
<td>.34**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM-MF Prac</td>
<td>-</td>
<td>-.56***</td>
<td>.54**</td>
<td>.39**</td>
<td>.55**</td>
<td>.41**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OFM Res</td>
<td>-.017</td>
<td>.018</td>
<td>-.053</td>
<td>-</td>
<td>.63***</td>
<td>.22**</td>
<td>.32**</td>
<td>.43**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OFM Champ</td>
<td>.221</td>
<td>.317</td>
<td>.109</td>
<td>-</td>
<td>-.29**</td>
<td>.37**</td>
<td>.38**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OFM Clin Sup</td>
<td>.049</td>
<td>.035</td>
<td>.321</td>
<td>-</td>
<td>-.009</td>
<td>.057</td>
<td>.002</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OFM Training</td>
<td>-.009</td>
<td>.090</td>
<td>.035</td>
<td>-</td>
<td>-.027</td>
<td>.24*</td>
<td>.082</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before Session</td>
<td>-</td>
<td>-.56***</td>
<td>.54**</td>
<td>.39**</td>
<td>.55**</td>
<td>.41**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treat Planning</td>
<td>-</td>
<td>-.63***</td>
<td>.22**</td>
<td>.32**</td>
<td>.43**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Client Feedback</td>
<td>-</td>
<td>-.65***</td>
<td>.47***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, ***p < .001.
uniquely explained 3.8%, 2.2%, 4.6% of the variance, respectively.

**Predictors for providing feedback of MLT scores to clients.** The model explained 32.6% of the variance in clinicians providing feedback of MLT scores to clients, adjusted $R^2 = .26, F (13, 143) = 5.31, p < .001$, a large effect ($f^2 = 0.48$). Significant predictors were therapeutic orientation, AM-MF practicality, and OFM clinical support, uniquely explaining 2.8%, 3.3%, 5.6% of the variance, respectively.

**Discussion**

The current study investigated the factors affecting the implementation of a routine outcome measure in naturalistic youth mental health settings, focusing on the use of MLT within MFS in headspace services across Australia. The first aim was to describe how much clinicians were using the measure and, as expected, clinicians were most likely to look at MLT before session, less likely use MLT in treatment planning, and least likely to provide feedback of MLT scores to clients. In most research, there is a discrepancy between the rates of use across MFS processes also seen in the current study (Ross et al., 2016). Providing quick and accessible outcome measurement data to clinicians is very important; however, clinicians have not always been shown to use the information when provided (De Jong et al., 2012). Several reasons have been identified for this disparity, including clinicians’ negative attitudes, anxiousness and discomfort, and lack of knowledge towards MFS (Ionita et al., 2016). The implementation of all MFS processes used together are most strongly associated with improved outcomes (Krägeloh et al., 2013). The use of MLT as a part of measures and their respective systems (Jensen-Doss et al., 2018). Participants identified MLT as mostly practical, moderately useful with treatment planning, and were somewhat positive about MLT’s clinical utility. These ratings are comparable or rated as more positive than clinicians views towards standardised progress measures generally (Jensen-Doss et al., 2018). This may be due to headspace’s design of MLT, which is easy to use, brief, sensitive to change, used session-by-session, and is presented instantly to the clinician before session in the form over a graph over time (Kwan et al., 2018). Notably, it was originally co-designed with the input of clinicians (and clients) and provides information that is meaningful to them. Development of measures that provide efficient information and are integrated into an accessible electronic system reduces the practical concerns that afflict other measures and their respective systems (Jensen-Doss & Hawley, 2010).

Participants were positive about the practical resources provided for MLT use and neutral about the organisation’s championing and training of MLT use. The only negative perceptions were about the level of clinical support provided for MLT use. The implementation of MLT as a part of the electronic data collection system used across headspace services is highly progressive. However, it was initially implemented for monitoring and evaluation purposes, rather than for clinical utility, which has only been more recently supported (Kwan et al., 2018; Kwan & Rickwood, 2020; Rickwood et al., 2018; Rickwood et al., 2015). Therefore, these results are unsurprising but reinforce the need for the organisation to focus on these neutral or negative perceptions of championing, training, and clinical support to support MLT use.

The second hypothesis, that clinicians would hold neutral or positive attitudes and perceptions towards MLT use in MFS, was supported with participants agreeing that there were benefits to MFS use and disagreeing that there were harmful effects, with large effect sizes consistent with previous research (Jensen-Doss et al., 2018). Participants identified MLT as mostly practical, moderately useful with treatment planning, and were somewhat positive about MLT’s clinical utility. These ratings are comparable or rated as more positive than clinicians views towards standardised progress measures generally (Jensen-Doss et al., 2018). This may be due to headspace’s design of MLT, which is easy to use, brief, sensitive to change, used session-by-session, and is presented instantly to the clinician before session in the form over a graph over time (Kwan et al., 2018). Notably, it was originally co-designed with the input of clinicians (and clients) and provides information that is meaningful to them. Development of measures that provide efficient information and are integrated into an accessible electronic system reduces the practical concerns that afflict other measures and their respective systems (Jensen-Doss & Hawley, 2010).

Participants were positive about the practical resources provided for MLT use and neutral about the organisation’s championing and training of MLT use. The only negative perceptions were about the level of clinical support provided for MLT use. The implementation of MLT as a part of the electronic data collection system used across headspace services is highly progressive. However, it was initially implemented for monitoring and evaluation purposes, rather than for clinical utility, which has only been more recently supported (Kwan et al., 2018; Kwan & Rickwood, 2020; Rickwood et al., 2018; Rickwood et al., 2015). Therefore, these results are unsurprising but reinforce the need for the organisation to focus on these neutral or negative perceptions of championing, training, and clinical support to support MLT use.

The third aim was to examine the strongest predictors of MLT use in MFS. Consistent with previous research, fewer years of clinical experience was shown to be a unique predictor of looking at MLT before session, but not for treatment planning or client feedback processes. It may be that clinicians with fewer years of experience have more recently graduated from programmes that are more likely to have incorporated routine progress monitoring in their clinical training (Ionita & Fitzpatrick, 2014).
Alternatively, they may be less confident in their clinical ability and may value the use of external and quantified feedback on clinical progress. CBT clinicians were more likely to provide feedback of MLT scores to clients when controlling for other factors, and it has been proposed that the structure of CBT is congruent with this type of monitoring and feedback (Jensen-Doss et al., 2018). While differences in years of experience and therapeutic orientations are not definitive predictors of treatment outcomes (Goldberg et al., 2016; Luborsky et al., 2002), clinicians with fewer years of experience are shown to have more anxiety, need more direction, and are still working to embed specific learnings. More experienced clinicians are more autonomous and move towards consistency between their worldview, self-concept, values, and therapeutic framework, which may not always match the standardised and regimented use of MFS (Rønnestad et al., 2019).

MLT practicality was the only unique predictor for all three MFS processes and accounted for a considerable 12.2% of the variance in clinicians looking at MLT before session. It is advised that creating or modifying outcome measures to make them more practical, or providing increased education around their practical benefits, helps increase their use in MFS (Jensen-Doss et al., 2018; Jensen-Doss & Hawley, 2010). This is particularly relevant for MLT as the results show that clinicians were very positive about MLT’s practical qualities. Unsurprisingly, positive attitudes toward MLT treatment planning properties were related to clinicians using MLT in treatment planning more often.

As hypothesised, organisational championing was a unique predictor of clinicians looking at MLT before session and using MLT in treatment planning, while organisational clinical support was a unique predictor of clinicians providing feedback of MLT scores to clients. Championing creates a positive culture which prioritises and values innovation of MFS use, which research suggests are one of the essential factors of MFS use (Gleacher et al., 2016). However, when the process of providing feedback to clients is examined separately, clinical support is critical, and there is a need for clinical guidance and expertise, not just championing. This is consistent with previous research that has identified training, clinical support, and a useful electronic system as the most important factors for the clinical processes of MFS use (Lyon et al., 2019).

**Practical Implications**

The practical implications of the current study inform supervisors and organisations of the areas of importance in supporting clinicians to use each MFS process. Workload resources and IT supports can help initiate the use of MFS (Brooks Holliday et al., 2020). These supports may reduce the barriers for clinicians with more years of clinical experience to look at measures before session. It may be helpful to target clinician attitudes towards a measure’s treatment planning properties to facilitate more use of measures in treatment planning. Supervisors should focus on clinical supports, such as individual clinical supervision and clinical team discussions, to increase clinicians providing feedback on measurement scores to clients. The use of MFS is an evidenced-based practice that is shown to be transatheoretical and transdiagnostic in nature (Boswell, 2019), and it may be helpful to involve non-CBT clinicians in discussions about how MFS can fit into their therapeutic framework.

At a service level, a focus on a measure’s practicality can help increase MFS use across processes, but particularly for clinicians looking at MLT before session. Organisational championing was identified as necessary for initiating MFS use in treatment planning, and targeted efforts should be focused in this area. Overall, these strategies target increasing the use of MFS to levels that may be more effective in producing clinical outcomes and formal feedback which has been shown to enhance the process of deliberate practice (Fortney et al., 2017; Macdonald & Mellor-Clark, 2015). Deliberate practice is the process of systematic effort to improve clinical performance, which has shown to produce more effective clinicians (Chow et al., 2015).

Importantly, benchmarks have recently been created for MLT in the form of clinically significant change indexes by age group and gender and expected change trajectories established by baseline MLT scores (Kwan & Rickwood, 2020). These benchmarks are yet to be incorporated within the headspace electronic data collection system, where they will provide additional clinical information for clinicians and clients. The creation of benchmarks, combined with health technology, allows clinicians to quickly identify potential blind spots and target areas during treatment (Chorpita et al., 2016). Given that clinicians have been shown to have limited knowledge around the use of routine outcome measures in predicting client deterioration, implementation of such benchmarks are vital (Bystedt et al., 2014). The addition of expected change trajectories and risk signals have been shown to enhance the treatment effects already seen in MFS (Delgadillo et al., 2018).

There is a large discrepancy between clinicians looking at MLT before session and using it in the other MFS processes. It is therefore recommended...
that formalised training, clinical supports, and treatment guidelines are created. This should include the evidence for the use of brief measures in MFS (Østergård et al., 2020), a guide on how clinical benchmarks can help inform treatment planning (Kwan & Rickwood, 2020), and a structured framework of how to discuss feedback with clients (Law & Wolpert, 2014). MLT consists of five items targeting areas of importance to young people, and these items can be individually tracked (Kwan et al., 2018; Moses & Claypool, 2018). This has shown to be a collaborative method of highlighting strengths and targeting problem areas for clients (Cross et al., 2015). Finally, there can be concerns from clinicians about MFS used as a performance management tool, and specific procedures should be provided for supervisors and organisations on how to appropriately utilise MFS to reduce these concerns (Law & Wolpert, 2014).

Limitations

The results of the current study need to be interpreted in light of its limitations. A fundamental limitation was the self-report nature of MLT use in MFS. Most studies only investigate the frequency of administration of routine measures, and only a few have explored the breakdown of MFS processes; specifically looking at measures before session, use of measures in treatment planning, and providing feedback of measure scores to clients (Ross et al., 2016). While this is an innovative way of exploring MFS use, the use of new technologies to track implementation may help increase the ability to measure MFS use objectively. Specifically, having IT systems that can log when clinicians access outcome measurement data could identify if clinicians access the data before session, during session to provide client feedback, and after session for treatment planning. The current research survey did not enquire about the clinicians’ caseload volume, or knowledge and experience with MFS and MLT, which have been shown to influence attitudes towards MFS and the amount of MFS use (Ionita & Fitzpatrick, 2014). Furthermore, headspace is an early intervention youth mental health service designed with a range of innovative service provisions for national use, and results of the current study may not be generalisable to other mental health care settings.

Another limitation was the possibility of a sample bias through voluntary survey recruitment. The survey was sent to headspace clinicians across Australia, and 17% who responded may have had more interest and experience with MLT and MFS. Such low response rates are typically seen with online surveys, however, and this level of response was to be expected (Manfreda et al., 2008; Morton et al., 2012). The results may not be representative of headspace clinicians generally, although the professional breakdown of this sample is broadly consistent with previous reports (Allott et al., 2019; Rickwood et al., 2015). Similar to the current study, a survey of headspace clinicians by Allott et al. (2019) consisted of predominately psychologists (> 40%), social workers (approximately 20%), and a range of nurses, counsellors, youth workers, occupational therapists and general practitioners (all under 10%, respectively). Additionally, on average they were also aged in their late thirties, had approximately nine years of clinical experience, were mostly postgraduate qualified, were mostly salaried staff, and just over a quarter were private practitioners.

Conclusion

The current paper provides an overview of clinician views on the utility of MLT and a summary of clinicians’ perceptions towards headspace supports in the use of MFS. This contributes to growing research supporting the need to implement routine measures within MFS for youth mental health care (Kodet et al., 2019; Mayworm et al., 2020). This study furthers the knowledge around the implementation of MFS, practically, identifying the factors for clinicians using the different MFS processes. MFS presents as a promising evidence-based platform to help clinicians target common factors of clinical change and enhance their deliberate practice (Boswell et al., 2015). As more efforts are made to understand the factors affecting the implementation of MFS, the more feasible it becomes for clinicians and organisations to utilise these systems.

Data and/or Code availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to restrictions that could compromise research participant privacy.

ORCID

BENJAMIN KWAN © http://orcid.org/0000-0001-9532-2990

DEBRA J. RICKWOOD © http://orcid.org/0000-0002-4227-0231

PATRICIA M. BROWN © http://orcid.org/0000-0002-6045-0068
References


Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment...


Chapter VII: Discussion

This final chapter presents a review of the aims of the research, a discussion of the main contributions of the four studies, and draws together the overall research findings for the thesis. The significance of the research regarding current knowledge and clinical practice is also considered, including a discussion of youth mental health outcome measures and MFS. Clinical practice implications are considered, including the use of electronic measurement systems, telemental health, and precision mental health care. The chapter also explores the use of MFS in clinician development, while considering the organisational implications of MFS use. The strengths and limitations of the research are identified, and directions for future research offered.

7.1 Summary of Aims and Research Findings

The primary goal of this research was to psychometrically test and examine factors affecting the implementation of a mental health outcome measure for young people aged 12 to 25 years, to be used within MFS. The research project targeted a developmental age group that has high rates of clinical deterioration, treatment dropout, and missed appointments, and who may be quite responsive to monitoring and feedback. The thesis comprised four studies aimed at identifying and addressing the gap in outcome measurements for young people, testing and optimising a session-by-session outcome measure (MLT), and investigating the factors for the implementation of MLT in MFS. Although each study had a distinct aim and different methodology, the research was designed to be progressive and sequential, starting with a review phase (Study 1), followed by a quantitative phase with young people (Studies 2 & 3), and subsequently a quantitative phase with clinicians (Study 4).

7.1.1 Study 1 – A systematic review of mental health outcome measures for young people aged 12 to 25 years. The aim of the first study was to identify appropriate
mental health outcome measures for young people aged 12 to 25 years, and this was achieved by conducting a systematic review of relevant research literature. Specifically, the aim of the review was to identify outcome measures that could be used for a broad range of mental health presentations, rather than be diagnostic specific, and that assessed mental health through global measures of cognition and emotion, functioning, quality of life, and multidimensional factors. The review also aimed to explore how the measures tracked change and in what populations and settings they had been used. Finally, the systematic review aimed to identify which measures were used routinely and used in MFS.

7.1.1.1 Age-appropriate measures. The systematic review identified 29 mental health outcome measures, reported in 184 articles, examining the change in mental health status for young people aged from 12 to 25 years. The identified outcome measures comprised eight measures of cognition and emotion, nine measures of functioning, six quality of life measures, and six multidimensional mental health measures. Five measures were deemed as possibly suitable for use across the whole 12 to 25 year age range, while also being applicable to a variety of clinical settings. It was noted that these five measures were originally validated with an adult population. Three measures were identified that were used with young people below 18 years, with an adult countermeasure used at follow-up. This reflects the design of outcome measures to match the traditional demarcation between CAMHS and AMHS within mental health care systems (McGorry et al., 2013).

7.1.1.2 Types of reporter measures. The systematic review identified four reporter options, with outcome measures completed by either the young person (self-report), parent/carer, teacher, or clinician, based on their perspective of the young person’s mental health. The systematic review identified 13 of the 29 measures as having more than one reporter on the measure. 20 measures were self-report by the young person, 16 measures were clinician-reported, 10 measures were parent-reported, and one measure was teacher-reported.
There were four outcome measures with options for a self-report, parent-report, and clinician-report.

Typically, the Clinical Global Impression scale-Improvement is clinician-reported, however, one article in the systematic review used the Clinical Global Impression scale-Improvement with all three reporters, and this study specifically aimed to explore differences between reporter types (Lewin, Peris, De Nadai, McCracken, & Piacentini, 2012). This study found that clinicians’ ratings of improvement were more conservative compared to ratings by the young person or parent. Numerous clinician-reported measures have been shown to lack interrater reliability and it has been suggested that other perspectives be included in clinical practice when using such measures (Schorre & Vandvik, 2004). Additionally, when a clinician can view other perspectives in MFS, positive treatment outcomes have been shown to increase (Bickman et al., 2011).

While different perspectives on outcome measures are critical for treatment, self-report measures are essential for young people (Bentley et al., 2019). Collecting measurement data from a young person’s perspective appropriately acknowledges their growing maturity and independence, while also supporting SDM to engage them in treatment (Arnett et al., 2014; Stiggelbout et al., 2012). During adolescence, young people increasingly self-initiate seeking help from professional services; self-initiation has been shown to increase linearly from 6%-7.8% of those aged 12 to 14 years, to 32.4%-41.2% of those aged 21 to 25 years (Rickwood, Mazzer, & Telford, 2015). Young people have reported that completing a measure can sometimes be more comfortable and less embarrassing than directly bringing up difficult issues with the clinician, while also finding it helpful to track their own progress (Stasiak et al., 2013). Fundamentally, young people have shown a preference to be better informed and in control of the measurement process, which can empower the young person to take charge of their own mental health (Stasiak et al., 2013).
The 10 parent-reported measures identified in the review were primarily used with young people aged under 18 years (Deighton et al., 2014). It is vital to appropriately introduce the measure and explain its utility with parents, as parents have reported difficulties with some of the terms and language used in measures (Moran, Kelesidi, Guglani, Davidson, & Ford, 2012). Parents who are provided with feedback on multiple outcome measure perspectives report this feedback to be helpful, as they gain a broader understanding of progress for the young person. The most recent models of youth-focused care recognise the critical role of family, and parent-reported measures may provide additional insights for clients up to 25 years of age (Rickwood et al., 2018).

Family inclusive practice is an approach to service delivery that empowers the family to take an active role in supporting a young person’s mental health care (Baker, Burgat, & Stavely, 2019). Family and parents can feel undervalued in the treatment process, with typical themes of boundaries around confidentiality and not being taken seriously (McCann, Lubman, & Clark, 2012). This approach suggests that parental involvement in a young person’s care is negotiated with the young person, with due regard to consent and confidentiality (Coates, 2016). This is particularly important during early adulthood, as family involvement in treatment usually changes or ceases in this transition phase from adolescence (Baker et al., 2019). Although help-seeking influence from family, for accessing in-person mental health care, decreases linearly with age, 15.5% of females and a larger 21.7% of males aged 21 to 25 years still report family as the primary influence (Rickwood, Mazzer, & Telford, 2015). Parent-reported measures used across the entire 12 to 25 years age range may need further consideration.

7.1.1.3 Methods for tracking change. A sub-aim of the review was to identify how the identified outcome measures tracked change over time. Seven measures utilised RCI when reporting treatment outcomes and, from these seven measures, five utilised CSI to analyse
client outcomes. RCI provides a more accurate rating of change as it takes into account the reliability of a measure and reduces the chance of measurement error (Wise, 2004), while CSI provides a statistical benchmark to identify if a client shifts from the dysfunctional population distribution to the functional distribution during treatment (Jacobson & Truax, 1991). Once these indexes are calculated for a measure based on the specific population being targeted, client outcomes can be categorised as deteriorated, unchanged, improved, and recovered. This provides more in-depth information to clinicians, organisations, and researchers compared with the more commonly used pre-post tests of statistical significance and effect size. With such low numbers of measures using RCI and CSI with young people, identified in the systematic review, it is recommended that such methods of tracking change are implemented in future studies (Donald & Carey, 2017).

7.1.1.4 Population groups. The systematic review identified three measures predominately used in non-clinical, community-based samples: the GHQ-12, KIDSCREEN-52, and Satisfaction with Life Scale. The remaining 26 measures were used primarily with clinical samples, and 10 measures were used with a mix of both clinical and non-clinical samples. In this systematic review, the GHQ-12 was found to be used only in non-clinical samples, however, a more recently published review (Bentley et al., 2019) found one study that used the GHQ-12 in a clinical setting (Tait, French, & Hulse, 2003). The study results identified that the GHQ-12 might be useful in detecting psychological distress in a range of community and clinical settings. This suggests that measures used primarily in non-clinical settings could be useful in clinical settings to detect psychological distress in young people, but would require further assessment. Diagnostically-generic measures that can be used across numerous settings, such as community-based, early intervention/sub-clinical, and clinical populations, can be very useful in examining patterns across people with differing mental health conditions (Burgess et al., 2015).
7.1.1.5 Measurement time frames. Another sub-aim of the review was to identify how often outcome measures were administered and the duration of treatment time in which measures were used to track change. Most of the measures identified in the review were used across a range of short to long-term time frames. Two measures primarily used in a long-term time frame were the KIDSCREEN-52 and GHQ-12. These measures were mainly used in non-clinical samples to explore changes in the general population over long periods.

One study captured in the review used the KIDSCREEN-52 to assess change on health-related quality of life in a sample of non-clinical young people after three years (Rajmil et al., 2009). The length of such measures limits their use routinely, however, as measures that take longer to complete are less likely to be used by clinicians. Conversely, long measures provide detailed assessment information that may be helpful in treatment planning and goal setting (Bentley et al., 2019). The systematic review identified only three studies that reported routine use of outcome measures, whereby the same measure was used at multiple time points during treatment. These measures used at multiple time points need to be brief and sensitive to change.

7.1.1.6 Measurement feedback systems. The systematic review identified two measures that were used in MFS, with both measures designed for children and adolescents under the age of 18 years. The first measure, the SxS-SDQ, has a self-report and a parent-report version. The study identified in the review aimed to quantify adherence to the use of SxS-SDQ in MFS and to provide some preliminary findings on its sensitivity to change (Hall et al., 2015). The study found that the family completed the SxS-SDQ in 57% of sessions. A statistically significant change was found on the clinician-rated Children’s Global Assessment Scale and the self-report SxS-SDQ post-treatment, while no significant changes were shown on the HoNOSCA and the parent-report SxS-SDQ. Furthermore, the study did
not have a comparison group, so it is unknown whether the significant changes were due to the effects of MFS and further evaluation is needed.

The second measure identified in the review was the Symptoms and Functioning Severity Scale, which has several versions: self-report, clinician-report, and parent-report. The study identified in the review utilised the different versions of the measure in MFS (Bickman et al., 2011). The methodology reported that the measurement data was fed back to clinicians, but the study did not specify if this information was fed back to the young person. Results showed that young people in the MFS group improved significantly faster than those in the control group, with effect sizes of .18, .24, and .27 for the self-report, clinician-report, and parent-report versions of the scale, respectively. Improvement was heightened when the clinician viewed feedback from multiple sources, with effect sizes increasing to .27 and .40 on the self-report and clinician-report versions, respectively, but effect sizes did not increase on the parent-report version of the scale.

7.1.1.7 Conclusion. Although the systematic review identified many measures used with young people aged 12 to 25 years, only five were used across this whole age range and were suitable for a range of treatment settings. Overall, the review found no measures explicitly designed for this age range. It was recommended that future research should focus on the development of mental health outcome measures designed specifically for young people aged 12 to 25 years. The review advocated for measures to be psychometrically tested across this age group for appropriate validity, reliability, and sensitivity to change. The design of measures is also important, with the review finding they need to be developmentally appropriate for the targeted age group, and meaningful to both clinicians and clients. Furthermore, it was suggested that measures be designed or modified for session-by-session use, to allow them to be used in MFS. The review only identified two measures used in MFS and recommended additional research in this area, targeting youth mental
health. There is growing evidence that MFS can provide additional clinical benefits in psychotherapy (Bickman et al., 2011), and this approach could be particularly beneficial for young people who are harder to engage and more likely to deteriorate during treatment (Gulliver et al., 2010; Warren et al., 2009).

### 7.1.2 Study 2 – Development and validation of MyLifeTracker: a routine outcome measure for youth mental health

The creation of MLT was undertaken by headspace National Youth Mental Health Foundation (headspace.org.au) outside the remit of this research project. As noted in the second study, the development of MLT specifically for use in headspace centres involved a series of steps including a literature review, consultation with clinical staff, clients, and the headspace youth reference group, and pilot testing. MLT was created to supplement other outcome measures used in headspace, including the self-reported K10 and the clinician-rated Social and Occupational Functioning Scale (Rickwood, Mazzer, Telford, et al., 2015). MLT was designed for young people aged 12 to 25 years, and this fills an important gap in outcome measures, as there were no measures specific for this age group identified in the systematic review that was conducted in the first study. Consistent with recommendations in the systematic review, MLT was also developed to be brief and diagnostically-generic, to be able to be used every session with young people with a wide range of presenting issues. MLT needed to be implemented in an electronic format and, currently, only a limited number of measures are freely available to be administered this way (Becker-Haimes et al., 2020). Therefore, for several reasons, a headspace-specific measure needed to be developed.

headspace is the Australian Government’s National Youth Mental Health Foundation, initiated in 2006 to provide early intervention in youth mental health for those aged 12 to 25 years (McGorry, 2007). The headspace data collection system, which includes the K10 and MLT, routinely collects information from clients. headspace clinicians are also required to
complete clinician-reported measures after each session, however, they are not required to use client-reported data in a mandated way and they can choose whether, and how, they integrate the data into their clinical practice. The system was initially implemented for quality improvement and program evaluation purposes (Rickwood et al., 2014), but has evolved to be increasingly used as a clinical support tool.

The second study aimed to examine the psychometric properties of MLT to determine its reliability, validity, and sensitivity to change. MLT was tested for the entire age range of 12 to 25 years, and both males and females. It was hypothesised that the five items constituting MLT would represent a single factor and that it would show good internal consistency. It was expected that MLT would have adequate test-retest reliability and show appropriate concurrent validity with standardised mental health outcome measures. Finally, it was hypothesised that MLT would display appropriate sensitivity to change over time.

The sample consisted of 75,893 adolescents and young adults between the ages of 12 and 25 years, who commenced their first episode of care at a headspace centre, for a mental health-related issue, between July 1, 2015 and March 31, 2017. A unidimensional factor structure was determined using exploratory factor analysis and subsequently confirmed through confirmatory factor analysis. The study provided support for clinicians and services to use MLT as a single total score to track mental health outcomes. Notably, the measure was shown to display measurement invariance across age and gender combinations. The reliability of MLT was demonstrated through good internal consistency and robust test-retest reliability. Evidence for the concurrent validity of MLT was found, as correlations between MLT and standardised outcome measures were as hypothesised. Specifically, MLT was strongly related to measures of psychological distress, life satisfaction, and wellbeing. Similar correlation coefficients were evident for various age groups and males and females, further demonstrating concurrent validity across age and gender for MLT.
The sensitivity to change of MLT was shown via significant improvements that were observed for males and females in all age groups. Importantly, these changes were moderate in effect size. An RCI of 18.27 points was estimated for MLT, and the proportion of clients reaching the RCI were generally comparable across MLT and the K10. The growth curve pattern demonstrated by MLT showed a steep increase during the early stages of therapy, subsequent flattening out, and then a slight increase again later in therapy. There were similar rates of change across age and gender groups. The rates of change shown by MLT were comparable to other studies using validated measures in youth-based settings (Warren et al., 2010).

This study offers evidence for the validity of MLT as a brief and practical measure to be used with adolescents and young adults. MLT can be completed efficiently by young people using a tablet or computer device. It can be incorporated into an electronic system that tracks session-by-session change, and produce instantaneous time-series charts for ease of use and interpretation. Such features provide support for MLT to be used in MFS, by providing clinicians with the ability to quickly look at measurement data before the session and bring visual graphs into the session. MLT targets meaningful areas of life for young people, rather than focusing on traditional measures of diagnostic symptoms. While the items are all tapping into a single dimension, the level of change within each item can identify areas where therapy is being effective or not. It was concluded that the use of MLT in MFS could help clinicians with treatment planning and therapeutic discussions in-session with clients, which has been shown to increase treatment outcomes (Østergård et al., 2020).

This study recommended MLT be further tested to support its use more broadly in other youth-based settings. It was also recommended that clinical benchmarks be created for MLT, which include norms or clinically significant change indexes. Such benchmarks would enhance the practical utility of the measure to aid in its interpretation. These provide
statistically derived standards for clinicians to compare clients’ progress against, which is critical when measures are used in MFS (Krägeloh et al., 2015). The study also identified the need for additional research exploring clinicians’ and young people’s views on MLT. More specifically, an investigation into the utility of MLT in MFS was suggested, to determine how clinicians and clients can use the measure to discuss outcomes and identify targeted interventions for areas of need.

7.1.3 Study 3 – A routine outcome measure for youth mental health: clinically interpreting MyLifeTracker. The third study aimed to identify statistically derived benchmarks for MLT within youth mental health services and provide a framework for how such benchmarks can be used in MFS. Initially, the study aimed to determine CSI for MLT that would identify developmentally appropriate thresholds for different age and gender groups. The study also sought to determine expected change trajectories and early warning signals for MLT, to provide benchmarks to help clinicians identify if a client is showing expected change over time in treatment, or whether a client is deteriorating. Lastly, the study intended to provide a guide and clinical examples of how clinicians can use the statistically derived benchmarks for MLT in their clinical practice.

Participants comprised both a clinical and a nationally representative community sample. Similar to Study 2, the clinical sample was between the age of 12 and 25 years, who commenced the first episode of care at a headspace centre between July 1, 2015 and March 31, 2017. However, only those who were deemed by their clinician to be at one of the following stages of mental illness were included in the analyses: mild to moderate general symptoms, sub-threshold diagnosis, threshold diagnosis, periods of remission, or ongoing severe symptoms. This resulted in a sample size of 63,840 adolescents and young adults for the clinical sample. The non-clinical group was a nationally representative community sample that consisted of 4,034 young people aged 12 to 25 years from across Australia. The
sampling was stratified to provide a near-even split between males and females, and across age groups.

The study found that CSI were distinctly different across age groups and gender, with male adolescents showing a higher threshold by 7-10 MLT points, compared to female adolescents. Young adult males also showed a higher threshold than young adult females, although this gap was smaller (2-4 MLT points). Overall, the largest difference between indexes was between males aged 12-14 years and females aged 18-21 years, with a difference of 18 MLT points. This clearly demonstrated the need for gender and age-specific CSI, to provide appropriate benchmarks responsive to the distinct developmental variances occurring during this rapidly changing time of life (Centofanti et al., 2019).

The results also showed that expected change trajectories followed a cubic pattern for MLT baseline scores below 50. This pattern of change using MLT is consistent with previous research that suggests there is a likelihood of more sudden change early in treatment, and then a deceleration as treatment progresses (Baldwin et al., 2009; Gaynor et al., 2003; Tang et al., 2002). Interestingly, rates of change were faster for young people scoring lower MLT baseline scores, compared to higher MLT baseline scores below 50, and these variances in rates of change are consistent with previous literature on youth outcomes (Cannon et al., 2010). There was an increasing, but slower, rate of change among baseline MLT scores between 51 and 70, whereas baseline MLT scores of 71 to 80 showed no change over time. As MLT scores approach the CSI, clients are closer to the functional distribution, and it is expected that less progress is likely to be seen. Baseline MLT scores of 81 to 100 showed an inverse cubic pattern to those baseline MLT scores under 50, with the MLT scores declining, and this is likely to be due to ceiling effects and/or regression to the mean, which is commonly seen with extreme measurement scores (Linden, 2013).
Lastly, the study provided two examples of how these newly created benchmarks and early warning signals could be presented within an electronic data collection system in the form of a graph over time. The first example presented a positive change directory, which showed the client following the expected change trajectory and finishing treatment above the CSI, based on their age and gender. The second example presented a client showing a negative change directory, dropping below the first early warning signal twice, before finishing treatment below the line, indicating deterioration. The examples are accompanied by a guide that summarises the clinical benchmarks for MLT, providing a brief explanation for each indicator of change and what it means in clinical practice. Such guidance is essential, given that clinicians have been shown to have limited knowledge around the use of routine outcome measures in predicting client deterioration (Bystedt et al., 2014).

The creation of benchmarks, combined with health technology, allows clinicians to quickly identify potential blind spots and target problem areas during treatment (Chorpita, Daleiden, & Bernstein, 2016). The addition of expected change trajectories and risk signals have been shown to enhance the treatment effects already seen in MFS (Delgadillo et al., 2018). MFS with such benchmarks can be beneficial for clinicians to inform clinical practice, provide feedback to clients, and support clinicians’ deliberate practice. Deliberate practice is a process of systematic effort to improve performance with the guidance of a supervisor, ongoing feedback relative to essential skills, and refinement and repetition of practice (Goodyear, Wampold, Tracey, & Lichtenberg, 2017). MFS aggregated data can also be utilised at an organisational level to inform service evaluation, quality improvement, and workforce development (Connors et al., 2020).

The paper concludes that further research is needed to determine MLT benchmarks for a range of settings for young people. Replication in other regions of the world focusing on the development of youth mental health systems, such as Canada, Ireland, the Netherlands,
and California, is warranted (McGorry, Trethowan, & Rickwood, 2019). The study used baseline scores as the predictor for growth curves, and it would be beneficial to explore other possible predictors of change, such as a client’s diagnosis and presenting issues. For example, a study examining the effectiveness of a substance abuse treatment found that while baseline mental health measures were a significant predictor of rates of change, employment status and baseline craving levels were also significant predictors (Crits-Christoph et al., 2015). Creating or modifying outcome measures to make them more practical and useful, and providing increased education around their clinical benefits, helps increase their use in MFS (Jensen-Doss et al., 2018; Jensen-Doss & Hawley, 2010). The study concludes by recommending future research that targets how to improve clinicians’ acceptability of MFS and how to enhance its implementation and effectiveness.

7.1.4 Study 4 – Factors affecting the implementation of an outcome measurement feedback system in youth mental health settings. The fourth study investigated the factors that affect the implementation of MFS in a naturalistic youth mental health setting, by examining the usage of MLT in headspace centres. Participants comprised 210 clinicians from 88 headspace centres across Australia. Participants were mostly female, psychologists, had a bachelor or master degree, used cognitive behavioural therapies, and were employed as salaried staff. Participants were recruited through an email invitation sent to 1176 clinicians and those interested in participating were directed to an online Qualtrics questionnaire. Data were collected between April 10, 2019 and July 31, 2019, and the response rate was 17.9%.

The first aim was to describe how much clinicians were using MLT and, as expected, clinicians were most likely to look at MLT before a session, less likely to use MLT in treatment planning, and least likely to provide feedback of MLT scores to clients. In the study, 13-24% of participants indicated no use of MLT across any of the three MFS processes. This is considerably lower than the proportion of clinicians who do not use
outcome measurements in other settings, which is shown to be up to 88% (Ionita & Fitzpatrick, 2014). The discrepancy between the rates of use across MFS processes shown in the current study is also evident in other research. For example, Ross et al. (2016) found that 83% of participants in their study administered the measures, 78% reviewed the measures, and 65% discussed the measurement results with their clients. The implementation of all MFS processes used together is most strongly associated with improved outcomes (Krägeloh et al., 2015). This suggests different active ingredients across the MFS processes and that an integrated approach is required for the full clinical benefits.

The second hypothesis was that clinicians would hold neutral or positive attitudes and perceptions towards MLT use in MFS. This was partly supported with participants agreeing that there were benefits to MFS use and disagreeing that there were harmful effects, with large effect sizes consistent with previous research (Jensen-Doss et al., 2018). Participants identified MLT as mostly practical, moderately useful with treatment planning, and were somewhat positive about MLT’s clinical utility. These ratings are comparable, or more positive, than clinicians’ views towards standardised progress measures generally (Jensen-Doss et al., 2018). Participants were positive about the practical resources provided for MLT use, and neutral about the organisation’s championing and training of MLT use. The only negative perceptions were about the level of clinical support provided for MLT use.

headspace’s electronic data collection system was originally implemented for research and service evaluation (Rickwood et al., 2018), and not as a clinical support tool. Unsurprisingly, the results reinforce the need for the organisation to focus on these neutral or negative perceptions of championing, training, and clinical support for MLT use.

The third aim was to examine the strongest predictors of MLT use in MFS. Fewer years of clinical experience was shown to be a unique predictor of looking at MLT before the session, and this could be explained by clinicians with fewer years of experience having more
recently had exposure to MFS in their clinical training (Ionita & Fitzpatrick, 2014). CBT clinicians were more likely to provide feedback of MLT scores to clients when controlling for other factors, and it has been proposed that the structure of CBT is congruent with this type of monitoring and feedback (Jensen-Doss et al., 2018). Unsurprisingly, positive attitudes toward MLT treatment planning properties were related to clinicians using MLT in treatment planning more often. MLT practicality was the only unique predictor for all three MFS processes, and it is advised that modifying outcome measures to make them more practical helps increase their use in MFS (Jensen-Doss et al., 2018; Jensen-Doss & Hawley, 2010).

As hypothesised, organisational championing was a unique predictor of clinicians looking at MLT before a session and using MLT in treatment planning, while organisational clinical support was a unique predictor of clinicians providing feedback of MLT scores to clients. Championing creates a positive culture that values the innovation of MFS use, which research suggests is one of the essential factors of MFS use (Gleacher et al., 2016). However, when the process of providing feedback to clients is examined separately, clinical support is critical and there is a need for clinical guidance and expertise, not just championing. Results of the current study are consistent with previous research that has identified training, clinical support, and a useful electronic system as the most important factors for supporting the clinical processes of MFS use (Lyon et al., 2019).

In conclusion, clinicians used MLT at different rates dependent on the MFS process, and each MFS process was further examined. The study provides an overview of clinician views on the utility of MLT and a summary of clinicians’ perceptions towards headspace supports in the use of MFS. This contributes to growing research supporting the need to implement routine measures within MFS for youth mental health care (Kodet, Reese, Duncan, & Bohanske, 2019; Mayworm et al., 2020). The study furthers the knowledge
around the implementation of MFS, identifying the factors that support clinicians to use the different MFS processes, which have rarely been separately explored.

7.1.5 Summary of studies. The four papers together produce findings with significant implications for mental health organisations and clinicians. The first paper provided an overview of broad outcome measures used with young people and this presents clinicians, organisations, and researchers with more informed information to identify the most appropriate measures dependent on the required need. However, the reviewed studies show a clear need for brief outcomes measures that can be used with a range of mental health presentations for young people aged 12 to 25 years. These are especially important in the context of reorientation of mental health systems to focus on young people who span adolescence and early adulthood, and who are diverse in their clinical presentation. MLT has been developed to address this gap and has now been thoroughly tested across all age groups (12-14, 15-17, 18-21, 22-25 years) and male and female genders, within primary care, community-based youth mental health settings. The addition of the clinical benchmarks provided in Study 3 allows MLT to become more clinically useful in its use in MFS. While increasing MLT’s usefulness can help increase its use in MFS, targeted implementation factors have now been identified for MLT use with each MFS process.

The importance of MLT being implemented into an electronic MFS has many benefits. Practically, MLT can be administered easily on a tablet device to young people, and there are instantaneous measurement graphs and comparisons to benchmarks that can be used in SDM (Hooke, Sng, Cunningham, & Page, 2018). This also allows MLT to be used in MFS when using telemental health, which has dramatically increased in 2020 due to the effects of the Coronavirus disease 2019 (COVID-19) pandemic (Douglas, Jensen-Doss, Ordorica, & Comer, 2020). Additionally, MLT integrated into a broader electronic data collection system can play a vital role in precision mental health care (Bickman, Lyon, & Wolpert, 2016).
There are broader implications for MLT use in MFS, which include its ability to support clinician development and organisational decisions. MLT can be used at an individual client level and also at an aggregated data level, to provide the necessary feedback for clinician deliberate practice (Chow, 2017). Appropriate training and clinical support are necessary for MLT to be used in MFS and for deliberate practice (Lewis et al., 2019). MLT provides *headspace* with another measure, among a suite of indicators, for use in service decision making and quality improvement (Rickwood, Mazzer, Telford, et al., 2015).

### 7.2 Measurement Feedback Systems with Young People

#### 7.2.1 The need for youth-specific, brief diagnostically-generic measures.

The development of valid, brief, and diagnostically-generic measures, such as MLT, provides a suitable option for MFS that is not available from longer measures or symptom-specific measures (Haroz et al., 2020). The brevity of such measures allows them to be used on a session-by-session basis and clinicians can receive the most up to date snapshot of a client’s mental health progress (Østergård et al., 2020). Brief measures are more likely to be acceptable to clients and used by clinicians, due to the reduced burden in their use. Broad measures targeting domains, such as quality of life or functioning, are transtheoretical and transdiagnostic in nature, meaning they can be utilised in an extensive range of settings with clients presenting with comorbid conditions, diverse diagnoses, and at different stages of illness (Scott & Lewis, 2015). With appropriate benchmarks, this can add more information into an assessment, by identifying if clients are more likely to present in the clinical or non-clinical population (Arroll et al., 2010).

Since the completion of the systematic review in this thesis, which was published in 2015, further measures have been tested for young people, showing the growing need and importance of this area (Lorentzen et al., 2020; O'Reilly et al., 2016). Increasingly, more brief
measures are being psychometrically tested and are found to be valid, reliable, and sensitive to change, while having fewer items (Haroz et al., 2020). For example, the YP-CORE (for those aged 11 to 16 years) and CORE-10 (for those aged 17 years and older) both have 10 items and are from the same suite of measures (Barkham, Mellor-Clark, & Stiles, 2015). Both measures have now been psychometrically tested for their respective age groups for young people accessing Jigsaw youth mental health services across Ireland (O'Reilly et al., 2016). The validation study included a break down into smaller age groups and gender, which is vital, however, there is still a significant disruption in changing measures at the age of 17 years, compared to MLT which covers the whole age range.

There are numerous measures that have been developed for one age group that have now been validated for use with other age groups. The GHQ-12, K10, and Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM), which were originally developed for adults, have now been tested with those under the age of 18 years (Centofanti et al., 2019; Lorentzen et al., 2020; Smout, 2019). There are fewer measures that have been tested specifically for young adults aged 18 to 25 years, which has been shown to be a distinct developmental age group (Klein et al., 2020; Rickwood, Mazzer, Telford, et al., 2015). The Columbia Impairment Scale-Youth Version, initially designed for young people under the age of 18 years, has now been validated for young people aged 15 to 24 years (Cleverley, Brennenstuhl, et al., 2020). Similarly, the SDQ, created for children and adolescents, has now been tested with young adults and comparable psychometric properties were shown (Brann et al., 2018).

Furthermore, since the systematic review in this thesis, studies have been completed on the use of measures, that were identified in the review, in MFS with young people (Tam & Ronan, 2017). The Ohio Scales and Y-OQ were identified in the systematic review, however, at that time were not used in MFS. A recent study has now tested the Y-OQ in MFS with
young people and found 64% improvement and 7.1% deterioration (Dyason et al., 2020). These results were compared to a study using the Y-OQ in TAU, which reported 57.1% improvement and 13.9% deterioration (Warren, Nelson, Burlingame, & Mondragon, 2012). There are increasingly more outcome measures being used in MFS with young people, although the evidence-base is still lacking compared to that for adult populations (Tam & Ronan, 2017).

The systematic review in this thesis focused only on mental health outcome measures, but another important focus for MFS is session-by-session therapeutic experience measures. A major focus of the therapeutic experience is the therapeutic alliance and some popular measures that target this include the Working Alliance Inventory (Horvath & Greenberg, 1989) and Therapeutic Alliance Quality Scale (Bickman et al., 2012). However, the therapeutic experience can also include client perceptions of treatment usefulness, service satisfaction, and appropriate expectations (Rickwood et al., 2017). Most measures have been designed to be used for research purposes, however, only a very few measures are created for routine use in clinical practice (Bull, Byrnes, Hettiarachchi, & Downes, 2019). The most commonly known therapeutic experience measure used in routine clinical practice in mental health is the PCOMS Session Rating Scale (Miller & Duncan, 2004). This measure is ultra-brief consisting of four items targeting the therapeutic relationship, treatment goals, consensus of treatment method, and overall experience. The therapeutic experience is associated with client treatment outcomes and, therefore, has been shown to be important when used in MFS (Bickman et al., 2012; Prescott, 2017).

MLT is an outcome measure that does not measure session satisfaction and therapeutic alliance, or have a partnered measure targeting the therapeutic experience, and this needs to be further explored. A brief measure routinely examining the therapeutic
experience has been recommended to be used across *headspace* centres and may serve as an essential addition to MLT being used in MFS (Rickwood et al., 2019).

**7.2.2 Implementation of measurement feedback systems.** The results of the studies described in this thesis have practical implications to help inform supervisors and organisations of the areas of importance in supporting clinicians to use each MFS process. Positive attitudes towards MLT overall, but especially related to its practicality and use in treatment planning, are crucial in its implementation in MFS. MLT can help provide efficient client-reported information in areas relevant to the young person, for discussion with the client and SDM (M. Slade, 2017). The addition of clinical benchmarks can help reduce clinician blind spots that are routinely revealed in naturalistic settings (Walfish et al., 2012). Studies 3 and 4 provide evidence for the utility and usefulness of MLT, which are essential for SDM. Specifically, SDM used with young people can promote the growing independence and maturation essential for this age group, by allowing them to be better informed and have more control over treatment options (Stasiak et al., 2013). This may help identify barriers in treatment progress, which may reduce the high deterioration and dropout rates shown in youth mental health care (Seidler et al., 2020; Warren et al., 2009).

The large discrepancy between MLT use in the different MFS processes, shown in Study 4, reveals a need for formalised supports within *headspace*, including training, clinical supports, and treatment guidelines. The guidelines need to cover the evidence for the use of brief measures, such as MLT within MFS (Østergård et al., 2020), a guide on clinical benchmarks, such as those identified in Study 3, a framework around decision making when using MFS (McKay & Coombs, 2012; McKay et al., 2012), and a structured outline of how to discuss feedback with clients (Law & Wolpert, 2014). While MFS allows for flexibility in its use, clinical examples such as those provided in Study 4 can provide some guidance on when treatment planning should be reviewed. Coombs et al. (2011) provide case studies of
the multiple uses of routine outcome measurement and MFS in Australia and New Zealand, which provides examples of the clinical utility of such systems. Jensen-Doss et al. (2020) have recently provided case examples using the Y-OQ with young people, on the use of MFS with evidence-based treatments, and using MFS to direct attention to common factors of treatment.

While most research, and the current thesis, has focused on MFS with individual clients, MFS can also be adapted and used in group work, family therapy, and couple’s therapy. The process can be more complicated, however, it allows the clinician to gather structured feedback from each client within a group session (Maeschalck & Barfknecht, 2017). The same outcome measures can be used in such settings, for example, the PCOMS stipulates the ORS can be used in group sessions to track individual wellbeing, however, the Group Session Rating Scale (Duncan & Miller, 2007) was created to assess the client’s experience of the group. The scale has been shown to be a reliable and valid way of capturing group engagement by each client, and asks clients to rate their experience of working with both the group clinician and the other group members (Quirk, Miller, Duncan, & Owen, 2013). Group MFS presents additional barriers to implementation, which include the additional time needed for the administration of measures and discussion of feedback with clients. Furthermore, some clients may not feel comfortable talking about measurement feedback scores in a group setting (Maeschalck & Barfknecht, 2017).

Lewis et al. (2019) completed a narrative review that helped inform a 10-point research agenda with practical implications critical for the implementation of MFS. The first two points involve harmonising the terms of the MFS processes and monitoring fidelity of their use. Point 5 recommends the need for brief and psychometrically tested measures, while Point 7 describes the use of electronic MFS in enhancing interpretability of measures. Point 6 identified timing of MFS use as an important agenda item, while Point 8 reported the need for
discrete evidence-based implementation strategies. The current thesis fulfils these six research agenda items, through the validation of MLT within an electronic system, determining clinical benchmarks to enhance MLT interpretability, identifying the frequency of MLT use across the MFS processes, and providing an in-depth investigation into the implementation factors affecting MLT use across the different MFS processes. Although, outside the scope of the current thesis, the four other research agenda points could be explored with MLT and through headspace in the future. This includes the use of MFS in exploring mechanisms for therapeutic change (Point 3), identifying active ingredients in MFS (Point 4), the implementation of evidence-based policies for MFS within organisations (Point 9), and the use of MFS in pay-for-performance models (Point 10).

7.3 Clinical Practice Implications

7.3.1 Electronic measurement feedback systems. There are increasingly more youth mental health outcome measures, that can be adapted into an electronic format, that are brief, accessible, psychometrically sound, and freely available (Becker-Haines et al., 2020). However, there are very few that have been tested to be sensitive to change and for routine progress monitoring. Some of the diagnostically-generic mental health measures that have been adapted into an electronic format include the K10, mainly used with adults, and the YP-CORE and SDQ, mainly used with children and adolescents. These measures are developed in line with the traditional age divide in most mental health care systems (McGorry et al., 2013). MLT addresses this gap by providing a measure that can be used across the 12 to 25 years age range, that has shown to be easily adapted into an electronic system.

One of the core benefits of electronic measurement systems is the instantaneous presentation of a client’s measurement score in a visual graph, with comparisons to clinical benchmarks (Overington & Ionita, 2012). Specifically, it has been shown that clinicians who
have access to graphs with expected change trajectories and risk signals have better client treatment outcomes, than those with access to the graph without these benchmarks (Delgadillo et al., 2018). Importantly, a study comparing clients’ perceptions of measurement feedback graphs, with and without expected treatment response trajectories, found that clients had a significant preference for the trajectories being included on the graphs (Hooke et al., 2018). Overall, clients found the trajectory graphs of more benefit in facilitating discussions with their clinician, which shows the importance of electronic systems being able to produce these expected treatment response curves.

A review completed by Lyon, Lewis, Boyd, Hendrix, and Liu (2016) explored characteristics and capabilities seen across electronic MFS. Four capabilities represented the core features of MFS: tracked standardised outcomes, feedback to clinicians and clients, visual graphs of measurement scores, and the provision of a range of measures. MLT used with the *headspace* electronic data collection system has these four capabilities and Study 4 identified MLT to be highly practical when utilised in an electronic system. Study 4 also showed that clinicians perceived that *headspace* provided a high level of practical resources for MLT use, which included MLT being implemented into an easily accessible system with a good design, while also providing appropriate hardware, such as computer and tablet devices to use MLT, and adequate technical assistance in the use of MLT in MFS.

### 7.3.2 Measurement feedback systems in telemental health

Telemental health is rapidly developing and there has been an increasing move toward more psychological and mental health interventions to be delivered via videoconferencing (Gloff, LeNoue, Novins, & Myers, 2015). Telemental health is defined as the delivery of mental health services via telecommunication modalities, including telephone-delivered therapy, videoconferencing, mental health applications, and internet-delivered programs (Reay, Looi, & Keightley, 2020). With the recent effects of the COVID-19 pandemic, this has been vital for mental health
service provision (Wind, Rijkeboer, Andersson, & Riper, 2020). When in-person supports were not available, clinicians and clients were more willing to utilise the digital format of supports and there was rapid uptake. At the start of April 2020, the proportion of *heads pace* services nationally that were delivered in-person reduced from 93% to 13% (headspace, 2020b). Most services were provided via telephone with an increase from 6% to 64%, while services provided through teleconferencing increased from 1% to 23%.

There are several barriers for clinicians using telemental health, which include its perceived impact on rapport building, the therapeutic alliance, privacy, and safety issues (Reay et al., 2020). Some clinicians believe that service provision using technology takes away from the quality of in-person connection (Orlowski et al., 2016). There is a reduced ability to identify non-verbal signals and many clinicians have little experience or knowledge with telemental health use (Reay et al., 2020). Most clinicians view telemental health as less effective compared to in-person care. In particular, implementation of telemental health with young people needs specific considerations, which include appropriate client space for sessions, how parents are integrated into sessions, the use of developmentally appropriate therapeutic aids, and appropriately trained staff (Gloff et al., 2015).

Recently, surveys were completed by young people and clinicians working with young people around the use of telemental health in *heads pace* services during the COVID-19 pandemic. With two-thirds of young people receiving services via telephone and just under a quarter via videoconferencing, most young people reported a positive experience (94%) and felt that it was suitable for their needs (78%) (headspace, 2020b). There were also several challenges identified, by both young people and clinicians, including interpersonal communication being more difficult, privacy issues when having a session from home, and frequent technical disruptions. Clinicians generally agreed that they could complete adequate risk assessment (69%), and half agreed that they could complete optimal therapeutic work.
utilising telemental health (headspace, 2020a). However, almost half of the clinicians in the survey felt that the therapeutic relationship and treatment progress were worse when using telemental health, while only about 10% felt the quality of the therapeutic experience improved. Nevertheless, the results show the acceptability for young people and clinicians to adapt to telemental health, although there are some concerns in maintaining the therapeutic connection and providing optimal treatment outcomes. It is important to note, however, that the transition to telemental health was rapid and unanticipated due to the pandemic. Yet, the results of the surveys show that the transition worked remarkably well for many clients and clinicians.

MFS integrated into telemental health services has been proposed as an ideal method for enhancing the virtual therapeutic process (Douglas et al., 2020). The framework for MFS when used in-person could be adapted to be used with videoconferencing, to increase engagement, therapeutic alliance, and treatment effectiveness (Chorpita et al., 2016). Several practical recommendations are suggested to increase the implementation efforts for MFS in telemental health (Douglas et al., 2020). A web-based MFS platform is recommended that provides the ability to send links by text message or email for clients to complete the measures. Clinicians should have the ability to share screens with the client to show and discuss measurement scores. Clinicians also should have multiple windows open, to simultaneously view the client, measurement feedback reports, and clinical support tools. Having multiple screens increases the feasibility of this process. Technological failures can significantly affect the therapeutic alliance and the effectiveness of the session (Lopez, Schwenk, Schneck, Griffin, & Mishkind, 2019), so steps should be taken to minimise possible pitfalls of technology, and a backup plan should be created, involving an alternative method of communication, which usually requires resorting to the phone.
Similar to using MFS in an in-person psychotherapy setting, clinicians should clearly introduce and explain the utility of the measures, and their importance in tracking change and SDM (Douglas et al., 2020). This is particularly important to help address the perceived reduced therapeutic alliance between clinicians and clients, specifically in telemental health services (Orlowski et al., 2016). The choice of appropriate measures in telemental health is essential, as clinicians have less ability to encourage young people to complete them. While not all MFS have clinical benchmarks and early warning signals, these are particularly crucial in telemental health, as these can help alert clinicians when clients are deteriorating or not on track, which may otherwise be missed in the online session format due to fewer non-verbal cues (Douglas et al., 2020).

 Appropriately, MLT has been explicitly designed with the qualities that are appropriate for telemental health. This includes its brief and easy to use nature, which reduces the burden for clients completing the measure every session. The items were developed in conjunction with young people and youth mental health clinicians, to ensure they are clinically meaningful. Furthermore, the items allow for targeted discussions in the virtual space to help with the therapeutic alliance. The clinical benchmarks for MLT are tailored for the young person, with adjustments by baseline MLT score, and age and gender. MLT is currently implemented in the headspace electronic data collection system, which can be sent to the young person via email or text message.

7.3.3 Precision mental health care. The MFS processes of initial assessment, ongoing monitoring, and individualised feedback are essential for precision mental health care. Precision mental health care is an intervention approach focusing on obtaining an accurate picture of a client’s mental health needs, preferences, and prognosis (Bickman, Lyon, et al., 2016). Such an approach provides tailored interventions using a data-driven selection of a specific treatment plan, based on the latest scientific evidence (Z. D. Cohen,
Kim, Van, Dekker, & Driessen, 2020). Precision mental health care has been proposed as the latest solution to enhance psychological treatments, using technological and methodological innovations. The research behind these systems can be categorised into two areas: pre-treatment decisions and adaptive decisions during treatment (Lutz, Rubel, Schwartz, Schilling, & Deisenhofer, 2019). More broadly, this approach has started to consider other factors shown to affect therapeutic outcomes, including clinician variability, therapy modality, therapeutic dose, and therapy ending type. MFS can provide client-reported information to make more informed pre-treatment decisions but, more importantly, can help monitor and discuss client feedback for appropriate adaptive decisions during treatment (Jensen-Doss et al., 2020).

Z. D. Cohen et al. (2020) developed algorithms to compare whether manualised CBT or psychodynamic therapy would be more effective for 167 clients with mild to moderate depression. The treatment comprised 16 individual sessions, completed within 22 weeks, by psychotherapists from outpatient psychiatric clinics in Amsterdam. The study found four factors that influenced the effectiveness of treatment: client depression severity, anxiety sensitivity, extraversion, and psychological treatment needs. It was theorised that clients with lower anxiety sensitivity, higher depression severity, higher levels of extraversion, and lower psychological treatment needs were better matched to CBT, and vice versa for psychodynamic therapy. The average post-treatment depression measurement score was 1.6 points lower (Cohen’s $d = 0.21$) for clients who received their indicated treatment compared to non-indicated. Among the 60% of clients with the strongest treatment indication, the final depression measurement score was 2.6 points lower (Cohen’s $d = 0.37$). Similar results have been found when comparing other indicated treatments, such as CBT versus person-centred counselling (Delgadillo & Gonzalez Salas Duhne, 2020), and suggest that treatment
outcomes could be improved if clients are matched to their optimal treatment model based on evidence-based algorithms.

Treatment outcomes can vary between clinicians, even when delivering the same psychotherapy model (Chow et al., 2015; D. Kim, Wampold, & Bolt, 2006). Given the significance of a clinician’s impact on outcomes, precision mental health care has started to develop prediction models indicating which clinicians are most effective for specific client presentations (Delgadillo & Gonzalez Salas Duhne, 2020). Kraus, Castonguay, Boswell, Nordberg, and Hayes (2011) found patterns of differential clinician effectiveness depending on a client’s primary presenting issues. In a sample of 696 clinicians in an adult outpatient setting, Kraus and colleagues found no clinicians were effective for treating all client diagnoses and 4% were not effective in treating any client presentations. An effective clinician was one whose average client reliably improved, while a harmful clinician was one whose average client reliably deteriorated. Some examples of the client primary presenting issues included depression, mania, sexual dysfunction, violence, and substance use, and correlations of clinician effectiveness between these presentations were relatively low. This suggests that clinician effectiveness is not a global construct and a clinician who is skilled in one diagnosis may be harmful in another. It was suggested that the results emphasise clinicians’ limited ability to question or evaluate their own professional abilities and the use of MFS is recommended to identify when clients are deteriorating, but also to identify clinician weaknesses in their professional development.

Another factor that impacts treatment effectiveness is the number of sessions provided, or the therapeutic dose, and this can be informed by MFS. A study investigating the dose-response effect of low-intensity and high-intensity CBT interventions found that the majority of low-intensity CBT clients reached RCI and CSI in seven sessions, while most high-intensity CBT clients achieved this in 14 sessions (Robinson, Kellett, & Delgadillo,
These results need to be taken in context, as those receiving high-intensity CBT are more likely to have more severe conditions that require more sessions. Specifically, clients with social anxiety disorder, obsessive compulsive disorder, or post-traumatic stress disorder had poor rates of improvement when utilising low-intensity CBT, and a better response to high-intensity CBT. Clients who accessed the full stepped-care pathway, utilising low-intensity CBT then being stepped up to high-intensity CBT, tended to have longer treatment length. At the same time, similar treatment responses were seen when clients were directly allocated to high-intensity CBT. This suggests that being able to allocate a client to the appropriate treatment stream quicker may provide more efficient outcomes. The use of MFS every session can quickly identify when progress is not occurring as expected and a client’s treatment plan can be reviewed to see if other interventions are required.

An ideal mental health system would implement a precision mental health care process that takes into account these numerous therapeutic factors (Delgadillo, Rubel, & Barkham, 2020). This process involves initially selecting the optimal treatment model to match the client’s presenting issues, such as CBT, psychodynamic therapy, or person-centred counselling (Z. D. Cohen et al., 2020). The second step is determining the optimal clinician match within that treatment modality, based on clinician effectiveness in a specific mental health domain (Kraus et al., 2011). The third step involves the use of MFS to ensure a client is progressing as expected, or whether the treatment planning needs to be adjusted (Shimokawa et al., 2010). This step would consider if the matched treatment model and clinician are meeting the needs of the client during treatment. It would also take into account if the number of sessions allocated or frequency of sessions are appropriate, and if the client needs to be stepped up into other higher intensity pathways (Robinson et al., 2020). Lastly, the use of MFS would allow clinicians to monitor if a client is likely to drop out of treatment and provide client feedback to explore treatment barriers (Lutz et al., 2020).
A new digital tool named the Innowell Platform aims to facilitate personalised and measurement-based care within one system (Iorfino et al., 2019). It combines several innovative and emerging practices, by including a range of measurement domains including social and occupational functioning, self-harm, suicidal thoughts and behaviours, alcohol or other substance misuse, physical health, and illness type, stage and trajectory (Hickie, Scott, et al., 2019). The system uses a clinical staging model and pathophysiological mechanisms to understand illness trajectories within a transdiagnostic framework. It allows for flexible use of a range of measures, depending on the presenting issues, and has the structured ability to inform assessment, client feedback, progress monitoring, and shared decision making (Iorfino et al., 2019). The combination of all this information can then inform evidence-based treatment planning and service provision (Hickie, Scott, et al., 2019).

The Innowell Platform has numerous, and generally lengthy, assessments and measures that require a substantial amount of time and burden for both clients and clinicians. Previous research has identified the need for measures for young people to be brief and easy to use (Hall et al., 2014). There have also been concerns from clinicians about electronic assessments taking too long for young people to complete and that the information received by clinicians can be overwhelming. Clinicians require summaries that are quick and easy to read, particularly when they receive them just prior to a consultation (Bradford et al., 2014). Systems, like the Innowell Platform, currently require a substantial commitment by clinicians and organisations, and the need to invest in substantial championing, training, and clinical support, as identified in Study 4. With the many measures and processes required for the Innowell Platform, there may be concerns about the practicality of such a system, which was identified in Study 4 as one of the most critical factors for the use of all MFS processes.

Several tools that could be used for precision mental health care across headspace have been tested and are currently being utilised within its electronic data collection system.
For example, MyAssessment was developed by *headspace* for young people aged 12 to 25 years, to help obtain assessment information quickly and easily (Bradford & Rickwood, 2015a). This is now the *headspace* electronic Holistic Assessment Tool, which is based on the HEADDS assessment (A. Parker, Hetrick, & Purcell, 2010; Rickwood et al., 2018). Similar to MLT, the Holistic Assessment Tool is integrated into *headspace*’s electronic data collection system, which can be sent out to the young person via email or text message prior to their initial assessment. Clinicians receive a brief summary of the young person’s mental health concerns before the session, equipping the clinician so that they can ask targeted questions based on the information received and identify areas of concern quickly for better informed decision making (Bradford & Rickwood, 2015a).

A number of other clinician-rated questionnaires are also included in the *headspace* electronic data collection system that increase precision mental health care, including the clinical staging model, diagnosis, Social and Occupational Functioning Scale, and treatment model provided (Cross, Hermens, & Hickie, 2016; Cross, Hermens, Scott, Salvador-Carulla, & Hickie, 2017). The inclusion of MLT provides a session-by-session measure to supplement the other intermittently administered self-report measures, such as the K10, and education, employment, and training status, to increase the use of precision mental health care across *headspace* (Cross, Scott, & Hickie, 2017). Precision mental health care with the inclusion of MFS aims to increase the effectiveness of treatments, by tailoring them to the relevant needs of clients, and can also help to develop more effective clinicians (Rousmaniere, Goodyear, Miller, & Wampold, 2017b).

### 7.4 Clinician Development Implications

There are particular challenges engaging young people in psychotherapy that include their high levels of perceived stigma accessing support and pervasive belief that seeking
assistance does not help (Gulliver et al., 2010). It is important that clinicians have a good understanding of change mechanisms for young people, to enable them to become more effective in clinical practice (Donald et al., 2018). They completed a longitudinal examination of clinicians’ professional development over time using client outcomes in naturalistic clinical settings. The study used 18 years of archival data from a counselling centre in a large American university, which consisted of 170 clinicians who had, on average, 4.73 years of data. They found that, on average, clinicians became less effective over time, although the effect was small. However, this should be taken within context as there was significant variation in clinicians’ trajectories over time. Interestingly, studies have consistently determined that 5% to 8% of the variance in client treatment outcomes were attributable to the clinician while adjusting for the initial severity of clients (Chow et al., 2015; D. Kim et al., 2006). Furthermore, clinician demographic and professional characteristics, such as age, gender, qualification, professional discipline, therapeutic orientation, and the size of their caseload, failed to predict client outcomes (Chow et al., 2015).

7.4.1 Deliberate practice. The Cycle of Excellence model proposes three components critical for superior clinician performance: determining baseline clinicians’ skills and strengths, utilising systematic and formal feedback, and engaging in deliberate practice (Rousmaniere et al., 2017b). Deliberate practice is a process of systematic effort utilising individualised professional development activities designed to improve aspects of a clinician’s performance through repetition and continual refinement (Ericsson, Krampe, & Tesch-Römer, 1993). It is proposed that the improvement of psychotherapy skills involves targets above the current abilities of the clinician, which require ongoing feedback and monitoring over an extended period (Goodyear et al., 2017). Engagement in deliberate practice is vital at all stages of a clinician’s career; it is critical for early skills acquisition,
while also being essential for more experienced clinicians to maintain performance (Chow et al., 2015).

A large scale study in the United Kingdom explored the effects of deliberate practice used by 69 clinicians from 45 organisations, on treatment outcomes from 4580 clients (Chow et al., 2015). The study used the CORE-10 to track client outcomes and utilised multilevel modelling analysis. The results found that the amount of deliberate practice undertaken by clinicians was shown to contribute to the differences between clinicians in client outcomes, with the most effective clinicians engaging in 2.8 times more deliberate practice than other clinicians. Deliberate practice was measured by how much clinicians engaged in 25 activities, which included engagement in supervision, attending training, reading clinical material, and reviewing cases.

There are four key components of deliberate practice: individualised learning objectives, immediate feedback, clinical supervision, and successive refinement (Chow, 2017). Individual learning objectives are regularly overlooked, and engagement in clinical supervision often lacks connection with the clinicians’ learning goals (Rousmaniere et al., 2017b). These learning goals are vital as they provide targets for clinicians to work towards, which takes them out of their comfort zone into their learning zone. Immediate client feedback helps identify any gaps during a client’s treatment and whether a clinician is utilising the appropriate skills to meet the client’s needs. Clinicians are shown to have a significant gap in identifying clients who are not on track and tend to underestimate levels of deterioration, which is proposed to be a self-assessment bias amongst some clinicians (Hatfield et al., 2010; Macdonald & Mellor-Clark, 2015).

Feedback about client progress can be about an individual client, or it can be an aggregate of outcomes over a clinician’s caseload (Chow, 2017). Using aggregated data for a
clinician can show a general level of effectiveness for specific client presentations, and this can allow clinicians to identify areas of strength and weakness. MLT used in MFS can be a prospective tool in supporting clinicians’ deliberate practice, and the provision of aggregated data can give the specialised feedback required for further development. It is proposed that such feedback is necessary, but not sufficient, in developing expertise. The next step builds on this feedback and comprises supervised successive refinement of skills (Tracey, Wampold, Lichtenberg, & Goodyear, 2014).

Repetition in the clinician’s practice can provide opportunities for gradual refinement and correction of errors with formative feedback (Bargmann, 2017). A large scale study of 275 clinicians conducted across 30 American universities, with 27,778 clients, found that effective clinicians at one point in time were also consistently effective across their caseload within that same setting (Owen et al., 2019). It was suggested that consistently effective clinicians had developed a refined approach for treating a specific clientele group. This emphasises the importance of being able to consistently perform well and is one of the main principles for clinical expertise (Miller, Hubble, Chow, & Seidel, 2013). Targeted feedback and deliberate practice are proposed as a method to reach this consistency, through the development of higher-order maps for case conceptualisation and treatment planning, which enable clinicians to act with greater precision (Owen et al., 2019). Lastly, clinical supervision should support the process of reviewing and redefining clinicians’ learning objectives, based on all feedback and observations of the clinicians’ practice (Bargmann, 2017).

7.4.2 Training and clinical support. Lack of training and clinical support for clinicians on the use of outcome measures and MFS is consistently identified as a barrier for their use (Lewis et al., 2019). Clinicians are more optimistic about MFS, find measures more practically useful, and are more likely to use outcome measures if they have received training (Hatfield & Ogles, 2004; Trauer, Callaly, & Herrman, 2009). Edbrooke-Childs, Wolpert, and
Deighton (2016) explored the use of ‘UPROMISE’ training on clinician attitudes towards MFS in CAMHS across England. The training was delivered as a one-day or three-day program, and included understanding and challenging barriers to MFS use, the usefulness of MFS, collaborative use of MFS with clients, and strategies for ongoing use. Both versions of the training increased clinicians’ positive attitudes and self-efficacy in MFS use. Unsurprisingly, clinicians who attended the three-day program had larger increases in levels of self-efficacy, compared to those who attended the one-day program.

There has been an increasing emphasis on accountability, competencies, and expertise in development for trainee or intern clinicians. It has been recommended that client outcomes are one of the criteria for professional development, and MFS can play an essential role for early-career clinicians (Goodyear, 2015). Clinical trainee programs could employ client feedback and progress as part of the learning outcomes within practicums and internships. A large study surveyed 605 doctoral psychology graduate trainees across the USA on their experience and attitudes towards the use of MFS in psychotherapy (Overington et al., 2015). The study found that 70% of surveyed trainees were aware of measures used in MFS and almost 60% had used them. When asked about their likelihood of future MFS use, 80% of users and 57% of non-users reported they would be likely to use MFS in the future. These proportions suggest that MFS have a reasonably high penetration in training settings and this may increase the future use of MFS. This is consistent with the results from Study 4, which identified fewer years of clinical experience as a unique predictor for looking at MLT before a session, suggesting that more recent trainees may have more exposure to MFS.

Another study surveyed 92 training clinic directors in the USA about their use of MFS in their psychology training clinics, as well as their current training practices in MFS for both trainees and supervisors (Peterson & Fagan, 2017). Results showed that most directors were providing their trainees with opportunities to use MFS (67.4%). Directors reported they were
also delivering targeted training on the use of MFS to both trainees (91.9%) and supervisors (55.7%). It was proposed that training supervisors increased their knowledge but also enhanced buy-in to implementing MFS in the training clinic. Supervisor training in MFS has been shown to improve supervisor attitudes to administering outcome measures and using feedback from measurement data, while also increasing levels of self-efficacy about the use of MFS in supervision (Fullerton et al., 2018). Given the importance of supervisors and champions in the implementation of MFS use, identified in this thesis, it recommended that training on MFS use be targeted to more supervisors.

Training alone is insufficient for changing and sustaining clinician behaviour, however, and it is recommended that ongoing supervision should be provided in MFS use (Connors et al., 2020). Even though there is usually an increase in MFS use after initial training, a decline is apparent with MFS use later in practice (Persons et al., 2016). A project that explored the implementation of MFS across 14 mental health workers in schools across a large Pacific Northwest district used a brief and efficient telephone consultation model for ongoing support (Lyon et al., 2019). The telephone consults supported troubleshooting technical issues and clinicians’ routine use of MFS processes. Results indicated that clinician MFS use increased in the subsequent week following the consultation call. It was suggested that the low intensive resource of a supervisory support call can help sustain MFS use.

A framework around supervision for MFS has been proposed, which can be used for both individual and group supervision (Bargmann, 2017). While individual supervision is mainly discussed in the literature, group supervision can enable shared ideas amongst colleagues, about how they integrate MFS in practice, and can also help address some scepticism and reluctance by some clinicians if others are seen using MFS (Lewis et al., 2019). The proposed framework for supervision on MFS consists of three steps. The first step is the clinician presenting a structured case presentation of a client, which is brief and based
on factual information. The idea of this is to prevent the clinician from imposing preconceived ideas on the case, allowing for more free reflection and new perspectives. The second step is examining the outcome measurement feedback graphs, which can indicate progress, no progress, or deterioration. Several factors then can be explored if a client is not progressing or is deteriorating, which include therapeutic alliance, treatment goals, therapy preferences, expectations for therapy, and if there is any missing information. The last step is reflection on all the available information and exploring as many ideas as possible. The reflections with the supervisor, or group, about treatment planning need to consider how far the client is into treatment and the client’s progress against the expected change trajectories (Bargmann, 2017).

The supervisor needs to balance two roles when providing supervision for MFS: an administrative role and a clinical role. The importance of these two roles is identified in Study 4, which shows the need for both championing and clinical support to increase the use of all MFS processes. The administrative role requires a focus on implementation efforts, such as how frequently the clinician is administering the measures, reviewing the data, and providing feedback to clients (Ross et al., 2016). The other vital aspect is ensuring the clinician’s fidelity to the MFS processes, which is shown to affect treatment outcomes (Krägeloh et al., 2015). The clinical role focuses on consulting with clinicians around case presentations and treatment planning ideas, as discussed in the framework above. Part of clinical supervision on MFS should be a culture of an open and transparent dialogue that sees errors as valued opportunities for learning (Bargmann, 2017). This means clinicians are more likely to discuss clients who are not progressing within supervision, which may reduce the blind spots and biases typically seen in clinical practice without MFS use (Z. J. Parker & Waller, 2015). It is also crucial that supervisors use MFS in their own practice and openly discuss their own clients who are not progressing, to model this culture of feedback.
There are some important concerns regarding the use of MFS systems outside the clinician-client relationship and clinical supervision. MFS can produce aggregated data for each clinician, which can include the number of active clients, the severity of clients, length of treatment, and dropout rates. Clinician-relative effect sizes can be calculated against the norm, which provide metrics for clinician effectiveness. Other methods of exploring clinician effectiveness can include identifying proportions of clients reaching CSI, progressing as expected, at-risk of treatment failure, or deteriorating (Bargmann, 2017). Such information can be used to support clinicians in deliberate practice, by exploring their strengths and weaknesses across their caseload. Guiding questions for supervisors to ask clinicians could include if the observed outcomes are what would be expected, if the treatment is being provided as it was intended, and if there is anything else that can account for discrepancies in outcomes (Connors et al., 2020). However, understandably there can be a considerable level of anxiety and uncertainty for the clinician using this data in supervision, with concerns how it may impact their employment (Law & Wolpert, 2014).

Precautions need to be taken when MFS aggregated data are used for supporting clinician development and for organisational decision making, which is discussed later. At an organisational level, frameworks should be in place so that MFS are introduced to clinicians as tools to support collaborative working and SDM, rather than for performance management (Wolpert, 2014). Specifically, it is recommended that clinician level data is utilised confidentially for internal workforce development and quality improvement to prevent unintended consequences related to misinterpretation of the data (Van Der Wees et al., 2014). Creating a culture of MFS use in clinician development needs to come from the top down (executives, managers, supervisors) and from the bottom up (clinicians) (Law & Wolpert, 2014). The supervisor and supervisee need to negotiate how they use MFS aggregated data together, to ensure collaboration, openness, respect, reflection, and skill development. While
analysing such data can be particularly sensitive for clinicians, it can be valuable in planning professional development initiatives for the clinician. The supervisor must aim to be as specific and precise in identifying areas for improvement, but also be non-judgmental and constructive (Bargmann, 2017).

7.5 Organisational Implications

7.5.1 Measurement feedback system implementation across different settings.

Clinicians are using MFS in a variety of treatment settings and client populations. Most research on MFS comes from large integrated health care systems and multi-service agencies, such as MLT used in 100+ headspace centres across Australia, which was the focus of the current thesis. This brings challenges for small practices and clinicians working in private practice by themselves (Lewis et al., 2019). These challenges can include no electronic health records or data collection systems, no administration or support staff, and lack of access to consultation from other clinicians. To overcome this, clinicians should implement self-management resources, which may include having a registry of clients who need a higher level of care, compared to those tracking as expected (Kroenke & Unutzer, 2017). There is a range of free, brief, and psychometrically tested hardcopy measures that could be utilised in MFS (Becker-Haimes et al., 2020), but the benefits of electronic web-based systems, that help track treatment change and can support other parts of client treatment, may outweigh the cost associated with these systems.

Some electronic measurement systems can be utilised free of cost for the basic features, however, most systems require a fee for various applications, which include licensing, technical support, software costs, client feedback reports, and benchmarking (Overington & Ionita, 2012). Additionally, the cost for the use of these systems usually include some form of training and user manuals, meaning there are initial one-off expenses
involved in the implementation of such systems. There are cost-effective packages designed for sole clinicians in private practice and for groups of clinicians within an organisation. However, the cost of MFS implementation can be one of the most significant barriers for some organisations and clinicians (Kotte et al., 2016).

Other factors also need to be considered with the implementation of electronic measurement systems in clinical practice. Firstly, measures included in these systems need to be appropriate for the target population group the clinician or organisation is working with. For example, the OQ-Analyst software is a web-based scoring system that utilises the OQ suite of measures, including the OQ-45 and the Y-OQ (www.oqmeasures.com). The OQ-45 has 45 items for those aged over 18 years and the Y-OQ has 64 items for those under 18 years. The PHQ-9 and GAD-7 are utilised in the Patient Case Management Information System (www.york.ac.uk/healthsciences/pc-mis). This electronic clinical record system is used in the Improving Access to Psychological Therapies program in the United Kingdom, which is a stepped-care model providing evidence-based, structured psychological interventions for depression and anxiety (Delgadillo et al., 2018). When choosing MFS, the needs of the target client group and the requirements of the organisation should be considered.

Clinicians can have privacy concerns when using external electronic systems and such a system should assure strict confidentiality. For example, the OQ-Analyst software complies with the Health Insurance Portability and Accountability Act of 1996 (Overington & Ionita, 2012). Other systems have additional features, such as an integrated note keeping system, which can reduce the number of systems used in practice (Delgadillo et al., 2018). Such features can reduce clinicians’ workload and burden in routine outcome measurement use, while enhancing the practicality of the system, which is one of the essential factors for MFS use identified in Study 4. MLT used across headspace centres does not currently have an
integrated note keeping system in their electronic measurement system, which would be ideal in such a large integrated health care agency.

In some large organisations, MFS are used by multiple clinicians providing a range of services, and these can be provided across different sites. The coordination of such processes can be complicated and it has been shown that different implementation efforts across sites utilising the same MFS can affect treatment outcomes (Bickman et al., 2011). Gelkopf et al. (2020) completed a systematic review on the implementation of MFS and found that most MFS were implemented in outpatient settings only (65%). The remaining studies were implemented in both outpatient and inpatient (24%), and 6% of the studies were in inpatient settings only (5% of studies had no available information). Most studies identified in this review targeted clients with a mix of diagnoses (56%), 16% had clients with anxiety and depressive disorders only, and 14% of studies targeted clients with serious mental illnesses (14% of studies had no available information). The implementation of MFS should be appropriately adapted when required for specific populations (Maeschalck & Barfknecht, 2017). The review identified that most studies indicated that measures were chosen based on their psychometric properties, which often focused on assessing functioning and symptoms, and only a few studies used measures of quality of life and recovery (Gelkopf et al., 2020). It was recommended that future projects use such measures in MFS that can be more meaningful to clients, and MLT provides a promising measure for such use.

7.5.2 Outcome measurement in organisational decision making. MFS are not generally mandated in organisations, however, routine outcome measurement is becoming mandated in some settings. For example, in Australia the National Outcomes and Casemix Collection were created in 2002, which specifies a set of standardised measures of symptoms and functioning, and a protocol for their collection at key points of care within public mental health services across all states and territories (Burgess et al., 2015). Data from these
measures are collected at a service level, then collated at a state/territory and national level to provide standardised reports of outcomes for different service types. This increases accountability and transparency of outcomes across agencies that can be compared for similarities and differences in service provision (Jensen-Doss et al., 2020; McKay, Coombs, & Duerden, 2014).

Aggregated data from routine outcome measures can be used at an organisational level to make data-informed decisions (Kilbourne et al., 2018). Specifically, the support of technology in MFS means that aggregated data can be readily accessible and up to date. Such information can help inform decisions for internal business functions, which may include appropriate allocation of cases to clinicians, identifying areas for service improvement, and targeting areas for workforce development (Jensen-Doss et al., 2020). Most diagnostically-generic outcome measurement systems can track new initiatives and programs, so that organisations can monitor implementation from the adoption and installation stages through to the maintenance and improvement stages. This means the same data collection system can be used in multiple programs for service evaluation, which reduces the costs and burden of having numerous methods of tracking outcomes (Kilbourne et al., 2018).

Pay-for-performance models, also known as value-based payment, are increasingly being recommended, particularly in the United States but also internationally, and these initiatives advocate to reward clinicians or organisations for outcome improvements (Kilbourne et al., 2018). These models provide reimbursement based on client reports of quality, rather than quantity of care, and usually require an electronic measurement system so information can be shared across stakeholders (Connors et al., 2020; Fortney et al., 2017). In the United Kingdom, these models are being implemented with local organisational goals being based on national quality improvement targets (Kilbourne et al., 2018). The push for these models has been from funding bodies and accreditation organisations demanding more
information about treatment outcomes. While funding bodies may be willing to support the
development of electronic measurement systems to enable value-based payment initiatives,
these systems need to be designed primarily to support client feedback and clinical decision
making in the first instance (Fortney et al., 2017).

As accreditation organisations begin to expect implementation of these models,
clinicians must have a sufficient understanding of how the data are used (Fortney et al.,
2017). When data are utilised in pay-for-performance models, there can be concerns from
clinicians or services, mainly when treatment progress is not as expected. It is recommended
that services leverage this information to review where the resources are being used, develop
actions plans for quality improvement, and identify top priorities (Connors et al., 2020).
Numerous factors can affect treatment outcomes, such as client motivation for treatment, the
severity of illness, sessions attended, and prognosis. This creates difficulty in reinforcing the
quality of care, while considering these factors that the clinician or organisation cannot
always control (VanLare, Blum, & Conway, 2012). Therefore, funding bodies must have a
proper understanding of how to interpret aggregated outcome data from MFS. This
knowledge should include awareness of the target population, what services are being
provided, and the benchmarks around expected change (Fortney et al., 2017).

The headspace electronic data collection system was initially implemented for quality
improvement and program evaluation purposes (Rickwood et al., 2014), and the inclusion of
MLT has supported the system to be used as a clinical support tool. MLT also provides
another validated measure of mental health for program evaluation. Electronic data collection
systems that support multiple uses can increase transparency in communication between
client and clinicians, and facilitate communication within the larger organisation (Jensen-
Doss et al., 2020). The need for accountability and quality improvement at a service level
must be balanced with clinicians being able to use MFS safely and effectively at a client
level. When implementing MFS, organisations should focus accountability on implementation efforts, such as targeting increased use of measures in treatment planning and providing measurement feedback to clients, but should proceed cautiously with expectations for clinical outcomes at a clinician level.

In 2003, the Australian Government’s Department of Health formed the consortium known as the Australian Mental Health Outcomes and Classification Network (Burgess et al., 2015). In 2012-2013 the network reviewed outcome measures used in community-managed organisations in Australia (AMHOCN & CMHA, 2013). They identified a shortlist of measures, that have been developed or used in Australia for mental health clients, that are brief and easy to use and had been psychometrically tested. However, it was advised that further research was needed to establish the suitability of these measures for this sector. Priority areas for future directions of routine outcome measures in Australia include an increased emphasis on clinical care at a clinician level, utilising measurement data at a service level, increased understanding of variations in processes and outcomes by funding bodies, and further research into the processes of clinical outcomes (National Mental Health Information Development Expert Advisory Panel, 2013). Furthermore, the National Mental Health Commission has recommended the integration of e-mental health with in-person services that are oriented to the client to promote self-managed care and SDM (Hickie, Davenport, et al., 2019). The results of the current thesis are consistent with these recommendations of routine outcome measurement in Australia.

7.6 Strengths and Limitations

This thesis by published works has a number of important strengths. A primary strength was the multiphase design that allowed multiple aims to be achieved over time. The use of these broad phases of research ensured that the most appropriate research approach
was adopted to achieve each aim, and the results of each study could support the next step. Overall, the thesis targets a developmental age group that has a high prevalence of mental health issues, low rates of help-seeking, and high rates of deterioration when in psychotherapy (Gore et al., 2011; Gulliver et al., 2010; Warren et al., 2009). The thesis contributes to the global research priorities for youth mental health, including the exploration of appropriate outcome measures, novel intervention methods, new technologies in treatment, and implementation strategies for innovative practices (Mei et al., 2020).

The use of a systematic review ensured that a comprehensive examination of the literature was completed utilising a robust and transparent framework (Moher, Liberati, Tetzlaff, & Altman, 2009). This approach allowed for a thorough exploration of broad mental health outcome measures for young people and the identification of gaps in the field. Specifically, it is the first review to examine outcome measures specifically across the 12-25 years age range. Prior reviews have reflected the traditional mental health service system demarcation by focusing on outcome measures used with either children and adolescents or with adults (AMHOCN & CMHA, 2013; Deighton et al., 2014).

Studies 2 and 3 had several strengths, the major one being the large sample size enabled by the large headspace centre network. This allowed for additional psychometric testing and calculations of CSI for smaller sub-samples across age groups and gender. This meant that distinct developmental differences were accounted for in this dynamic period with young people (Van Droogenbroeck et al., 2018). Many outcome measure validation studies cannot conduct robust psychometric testing due to inadequate sample size (Bentley et al., 2019).

Generally, sensitivity to change is not assessed when measures are psychometrically tested (Bentley et al., 2019). Studies 2 and 3 provided several indicators of change during
treatment, which is another strength of the thesis. The average magnitude of pre- and post-treatment changes were assessed, and this was followed with calculations of RCI, CSI, and expected change trajectories, which provide further depth in reporting changes in treatment (Donald & Carey, 2017). The proportion of changes using RCI on MLT were compared to the well-validated K10 measure. Importantly, the large national community survey conducted by headspace provided critical general community comparison data for the calculation of CSI. The modelling of these benchmarks is based on previous research that is evidence-based and statistically derived (Centofanti et al., 2019; Finch et al., 2001). Providing such a range of change indicators increases the interpretability of the measures’ ability to track change.

Study 4 provided clinicians’ perspectives on MLT and MFS, and clinicians’ views are essential in scale development and implementation science (Boateng, Neilands, Frongillo, Melgar-Quiñonez, & Young, 2018; Lewis et al., 2019). This again drew on the large headspace centre network to yield a relatively large sample of clinicians. Another notable strength in Study 4 was the breakdown of the MFS processes, which provided additional depth in practical implications for MFS use. There is limited research regarding the different MFS processes and these should be further investigated (Brooks Holliday et al., 2019; Ross et al., 2016).

The current thesis also has some notable limitations. The first is generalisability. The studies in this thesis were conducted across headspace centres, which are primary care mental health centres that support young people aged 12 to 25 years in Australia. This is a unique and broad population from early intervention youth mental health settings, and results may not be generalisable to other settings or other countries. Psychometric properties and clinical benchmarks for MLT would need further research for young people attending specialist or tertiary services. Replication in other regions of the world focusing on the development of youth mental health systems, such as Canada, Ireland, the Netherlands, and California, is also
warranted (McGorry et al., 2019). However, the thesis provides a structured example of how such measures can be implemented into MFS, that follows the latest recommendations (Lewis et al., 2019).

While a thorough search strategy was employed in Study 1, it is possible that relevant measures were missed due to the strict exclusion criteria. The review only included studies in the English language, and ‘grey’ literature and unpublished studies were not included. These criteria were employed to make the review more manageable, nevertheless, some measures may have been overlooked. Furthermore, more work has been undertaken since the publication of the systematic review.

While Studies 2 and 3 had major strengths due to the large sample size, some analyses were not conducted. Configural and metric invariance was tested across age groups and gender for MLT, however, further invariance tests were not conducted. While measurement invariance was demonstrated across age groups and gender, the number of observations may have masked variability. Specifically, there may be important differences in the factor structure of other population groups, such as those who identify as Aboriginal and Torres Strait Islander, culturally and linguistically diverse, and LGBTIQ+. Measurement invariance tests across these groups could be tested in future research.

MLT was used in the naturalistic clinical setting of headspace and sensitivity to change for MLT could not be compared with a control group, which is a limitation of the thesis. Furthermore, expected change trajectories could only be estimated for treatment lengths of up to 26 weeks, as headspace does not generally provide long-term therapy. This limits its use in MFS for longer-term psychotherapy. There is limited research of MFS effectiveness in longer-term settings and this is work that is needed (De Jong et al., 2014). Furthermore, the study is limited to expected change trajectories based on baseline MLT
scores only. Other factors have also been shown to be important in predicting average rates of change in treatment, such as substance craving levels and employment status (Crits-Christoph et al., 2015).

In Study 4, clinicians estimated their own amount of MLT use across the three MFS processes. Measuring adherence to MFS use in such ways has been criticised as biased, compared to practice observation, although multiple studies have found significant correlations between clinician self-report and observational assessments (Lewis et al., 2019). Another limitation of Study 4 was the possibility of a sample bias through voluntary survey recruitment and the low response rate. While the sample size was large enough to conduct the appropriate quantitative analyses (Tabachnick & Fidell, 2007), the participants may have been those with more favourable views of MLT and MFS. Nevertheless, the professional breakdown of Study 4’s sample was broadly consistent with other research reports from headspace, suggesting that there may have been sufficient representation (Allott et al., 2019; Rickwood, Telford, et al., 2015).

7.7 Future Research Directions

7.7.1 Measurement feedback systems with young people. Since the systematic review in this thesis, there has been some growth in the development and testing of outcome measures being used with young people aged 12 to 25 years. There is also still a large gap in these measures being appropriately tested in a range of clinical settings (Bentley et al., 2019). Similar to the testing completed with MLT, being able to test other measures in large samples for differences across smaller age groups and gender within this 12 to 25 age range is essential, as the current study has shown distinct developmental disparities. Furthermore, this can result in determining more developmentally appropriate clinical cut-offs for use by clinicians in day-to-day practice.
While MLT presents as a suitable measure for clinical use with young people, other appropriately validated outcome measures from different mental health domains also need to be created or adapted for session-by-session use. A number of brief measures have shown lack of sound methodological testing (Bentley et al., 2019). Intuitively, further research should explore the work on appropriate benchmarking of different session-by-session outcome measures across different settings. This should include RCI, CSI, expected change trajectories, and early warning signals for MFS, and the current thesis provides a framework for the creation of such benchmarks. It is recommended that calculations for such benchmarks are made more transparent and straightforward, as such methods are somewhat underutilised (Donald & Carey, 2017). Being able to show the appropriate statistically derived benchmarks provides clinicians with more confidence in the data and allows for better interpretation of measurement scores (Delgadillo et al., 2017).

With the rapid expansion of electronic measures and MFS, further exploration is required around client attitudes regarding this format of tracking change. MLT was co-designed with young people and clinicians, but further work is needed for young people to be included in the participation and development of electronic MFS, and the implementation of the MFS processes in general, particularly how it is used in client feedback and shared decision making. There is some research suggesting that young people find electronic measures more engaging and that they increase levels of use (Hall et al., 2014). Further research is needed to ensure the validity of electronic measures, now that more measures are adapted to be used in such ways (Chung & Buchanan, 2019). While there is some research exploring adolescents’ views on MFS (Philippa Moran et al., 2012; Stasiak et al., 2013), there needs be further exploration across the entire 12 to 25 years age range.

There is limited research into MFS in youth settings, and more in-depth research is required to develop a better understanding of the active ingredients in MFS and the
moderators to its beneficial effects. Similar to the adult literature, there needs to be more research on MFS effects used in different mental health settings, and across different psychotherapies and mental health presentations. Additionally, most of the research exploring youth MFS utilises the ORS, with only a few other measures shown to be effective in MFS (Tam & Ronan, 2017). More outcome measures for use in MFS need to be tested, which can assess whether other domains of mental health can be useful in such systems, while also being co-designed with clients and clinicians to be meaningful.

Other factors that are generally lacking in most MFS literature include the details of client dropout rates, the number of sessions provided during treatment, how much training on MFS was provided, and fidelity to use of MFS processes (Gondek, Edbrooke-Childs, Fink, Deighton, & Wolpert, 2016). This makes the comparison between the effectiveness of different MFS difficult, as these additional factors can differ between studies. There need to be further trials examining the effectiveness of MFS processes and how the different methods of MFS can be administered, such as the intensity and frequency of MFS use (van Sonsbeek, Hutschemaekers, Veerman, & Tiemens, 2014). This type of assessment would help clarify the most beneficial schedule of outcome feedback and the active ingredients to treatment effectiveness. Methodologically, most studies exploring youth MFS have limitations, including small sample sizes, lack of statistical power, incomplete data, and minimal procedural details (Tam & Ronan, 2017). Thus, researchers should target larger samples that can accommodate missing data and provide more robust and transparent reporting procedures.

The research investigating the implementation of MFS has expanded, although there is still limited literature focusing on young people (Jensen-Doss et al., 2020). There is still further research that would benefit MFS implementation science, including further investigation into the factors affecting the use of different MFS processes. While there is
some research exploring the breakdown of MFS processes (Brooks Holliday et al., 2019; Ross et al., 2016), further work is needed to provide more targeted and individualised implementation support for clinicians. This would help provide further clarity on the feasibility of MFS, particularly in youth mental health settings. There is evidence showing improved treatment outcomes when MFS implementation efforts are appropriate (Bickman, Douglas, et al., 2016). The mix between implementation efforts and treatment outcomes should be further assessed, to identify areas that produce the most effective or efficient results.

Clinician fidelity to MFS processes will increasingly be an issue as these systems become more widespread, and it is recommended that there is further development of guides for appropriate use of MFS. When all processes of MFS are used appropriately, there are more substantial treatment effects (Krägeloh et al., 2015). Some guides specific to certain measures and MFS have been developed, and also some more general guides on how to appropriately use client feedback in treatment planning (Law & Wolpert, 2014). However, there needs to be an expansion in this area to provide more specific guides for different treatment approaches with the use of MFS. To date, CBT has been utilised with MFS and has been shown to provide higher efficiency of treatment outcome, compared to CBT without MFS (Janse et al., 2017). Structured guides and frameworks could be created for using MFS with other therapeutic orientations. It is argued that MFS are transtheoretical and transdiagnostic evidence-based practice, however, MFS may need to be adapted for different therapeutic approaches, specific population groups, and different settings, including the use of different measures (Boswell, 2019).

Further areas of research for MLT, specifically, would initially be testing the measure in other youth mental health settings to be more generically useful across settings. MLT RCI and CSI could also be determined for other settings, to identify clinical benchmarks for that
target population. This thesis did not investigate young people’s experience using MLT and their views on the MFS process, or the effectiveness of MLT used in MFS, which could be conducted in future research. Research exploring MLT effectiveness should target clinicians’ caseload volume, knowledge and experience with MFS, and fidelity to MFS processes. The studies contained in this thesis are purely quantitative, and a qualitative approach in future research with young people and clinicians would be helpful, specifically to explore the different MFS processes and how they impact the experience of therapy. Qualitative methodologies can help capture the complex process of psychotherapy in thorough detail, which can help advance this area of clinical practice (Donald et al., 2014).

7.7.2 Clinical practice implications. Electronic measurement systems have allowed the possibility of MFS to be integrated into telemental health and precision mental health care. The rapid expansion of telemental health use during the COVID-19 pandemic has meant that clinicians are looking for better ways to engage clients within this format of therapy. MFS used in telemental health has been proposed as one solution to enhancing the therapeutic relationship (Douglas et al., 2020). Several electronic MFS are already designed with the ability to be used in telemental health, including MLT used within headspace’s electronic data collection system. While there are similarities to MFS use in both in-person therapy and telemental health, there are some additional steps to consider with MFS used in telemental health. Further research should be conducted investigating the implementation of MFS in telemental health and also the effectiveness of MFS within the online format.

The number of psychotherapy models has grown rapidly in modern times, however, the effectiveness of psychotherapy has not shown dramatic improvements (Miller et al., 2013). Precision mental health care is proposed as a solution for enhancing treatment outcomes in mental health. Future research should focus on the use of MFS in supporting precision mental health care with young people (Lutz et al., 2020). For example, a precision
mental health care framework could be utilised in conjunction with the *headspace* electronic data collection system, to be tested across centres. This could help further reduce the high deterioration rates and treatment drop out seen with this developmentally vulnerable population group during mental health treatment (Seidler et al., 2020; Warren et al., 2009).

7.7.3 Clinician development implications. Measurement data is proposed to be essential feedback of client progress and outcomes to clinicians, for the development of clinical skills (Tracey et al., 2014). However, even when clinicians receive such objective information they do not always utilise it appropriately to inform deliberate practice (Ross et al., 2016). Therefore, it is recommended that future research explore the effectiveness of MFS in increasing deliberate practice (Overington et al., 2015). This should include specific training frameworks and guidelines for the use of MFS and its use in clinical supervision. Organisations such as *headspace* could explore how MLT and other outcome measures can help inform deliberate practice.

It is recommended that the use of MFS and deliberate practice be incorporated into credentialing across all professional bodies and organisations that utilise psychotherapy. Accurately assessing competencies amongst clinicians is challenging and is often rated by supervisors with varying degrees of validity and reliability (Rousmaniere, Goodyear, Miller, & Wampold, 2017a). Most professional disciplines require ongoing supervision, but a critical inclusion to this work is the continued skill development and exploration of objective outcomes. Future research could focus on how well professional bodies and organisations can track effectiveness through outcomes, and how they can use this to support deliberate practice and clinician development (Goodyear et al., 2017). Although tracking treatment outcomes is only one component in clinician development, it can make a valuable contribution to improving professional practice.
7.7.4 Organisational implications. There is now clear evidence for the need for a structured framework for the implementation of MFS for best treatment outcomes (Lewis et al., 2019). To ensure clinician fidelity to the MFS processes, it is recommended that well-defined guidelines are created that include the determination of appropriate administration of measures, how client feedback should be provided, and coordinated response on how to execute treatment changes and follow-up when needed (Law & Wolpert, 2014). Results from Study 4 help inform such strategies for MFS use across the different processes in youth mental health settings, particularly supporting outcome measure practicality, championing, and clinical support.

Frameworks and guidance for service managers are also needed to support the use of data-informed decisions at a service level (Kilbourne et al., 2018). Future research could explore the effectiveness of structured decision making using routine outcome measurement at an agency level. This could include investigating models such as the pay-for-performance approach, to see if they increase the effectiveness of the service and clinicians' attitudes towards MFS in practice (Fortney et al., 2017). Implementation of such systems would require service managers to have a high level of knowledge around the interpretation of data and change processes within mental health treatment (VanLare et al., 2012).

7.8 Conclusion

This thesis aimed to further knowledge regarding mental health outcome measures used in MFS for young people and implementation into clinical practice. The results have made a significant contribution to the literature. The systematic review was the first to examine mental health outcome measures specifically for the 12 to 25 years age range. This was urgently needed to match the shift in mental health service provision to this developmentally vulnerable age range (McGorry et al., 2013). The review revealed a major
gap in diagnostically-generic mental health outcome measures designed specifically for this age group, and a particular lack of session-by-session measures that were appropriate for MFS.

The review justified the need for a new outcome measure that would address this gap. Consequently, the thesis provided a study on the validation of MLT, a routine outcome measure for young people aged 12 to 25 years. MLT is brief, easy to use, clinically meaningful, and sensitive to change, and has been integrated into an electronic MFS. MLT was thoroughly psychometrically tested with a large sample size that is not commonly seen with most measures (Bentley et al., 2019). It is now essential for measures used in MFS to have clinical benchmarks and the third study in the thesis provided CSI and expected change trajectories for MLT. While many measures are still lacking in such benchmarks, the third paper provides a framework for the development of CSI and expected change trajectories that can account for age, gender, and severity of mental health issues. The addition of a practical guide on how to interpret MLT benchmarks added to the limited practical resources found in the literature (Law & Wolpert, 2014).

Lastly, the thesis identified predictors for clinicians’ MLT use in MFS in youth mental health settings. Part of this project provided evidence for acceptability and usability of MLT in clinical practice, and this is essential in the scale development process, but not always demonstrated (AMHOCN & CMHA, 2013; Boateng et al., 2018). Furthermore, there is limited research in the literature exploring the implementation factors for the different MFS processes, and the current thesis provided further knowledge into this area. The evidence-base is still lacking in the implementation of MFS with young people, and this thesis adds to this vital area of research.
For the youth mental health system, there is further need for innovation in treatment models to seek progress in mental health care. Australia is leading the way in mental health care models for young people aged 12 to 25 years, with the development of designated primary care services such as *headspace*, and similar models are being implemented internationally (McGorry et al., 2014). These models acknowledge the biopsychosocial development and complexities faced by young people as they transition into adulthood, and the burden of mental health issues experienced by this age group. The implementation of MFS within youth mental health care is essential for progressing better outcomes for young people. With the high rates of mental health issues amongst young people and the ongoing impact this has into adulthood, innovations in psychotherapy targeting this age group are urgently needed.
References


Athay, M. M., Riemer, M., & Bickman, L. (2012). The symptoms and functioning severity scale (SFSS): Psychometric evaluation and discrepancies among youth, caregiver, and


Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quíñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and
doi:10.3389/fpubh.2018.00149

doi:10.1080/10503307.2019.1630780

doi:[https://doi.org/10.1016/S2352-4642(17)30007-X](https://doi.org/10.1016/S2352-4642(17)30007-X)


organizational goals. *Administration and Policy in Mental Health.*

doi:10.1007/s10488-020-01063-w


viewed favorably but rarely used in practice. *Administration and Policy in Mental Health, 45*(1), 48-61. doi:10.1007/s10488-016-0763-0


Mayworm, A. M., Kelly, B. M., Duong, M. T., & Lyon, A. R. (2020). Middle and high school student perspectives on digitally-delivered mental health assessments and


doi:10.1037/ccp0000437


doi:10.1016/j.brat.2014.12.010

doi:http://dx.doi.org/10.1016/S0140-6736(07)60368-7


doi:10.1016/j.brat.2015.11.004


doi:10.1186/1477-7525-3-76


### PRISMA 2009 Checklist

<table>
<thead>
<tr>
<th>Section/topic</th>
<th>#</th>
<th>Checklist item</th>
<th>Reported on page #</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TITLE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>1</td>
<td>Identify the report as a systematic review, meta-analysis, or both.</td>
<td></td>
</tr>
<tr>
<td><strong>ABSTRACT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structured summary</td>
<td>2</td>
<td>Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.</td>
<td>2-3</td>
</tr>
<tr>
<td><strong>INTRODUCTION</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td>3</td>
<td>Describe the rationale for the review in the context of what is already known.</td>
<td>3-7</td>
</tr>
<tr>
<td>Objectives</td>
<td>4</td>
<td>Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).</td>
<td>7-8</td>
</tr>
<tr>
<td><strong>METHODS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol and registration</td>
<td>5</td>
<td>Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.</td>
<td>NA</td>
</tr>
<tr>
<td>Eligibility criteria</td>
<td>6</td>
<td>Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.</td>
<td>8-9</td>
</tr>
<tr>
<td>Information sources</td>
<td>7</td>
<td>Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.</td>
<td>8</td>
</tr>
<tr>
<td>Search</td>
<td>8</td>
<td>Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.</td>
<td>8</td>
</tr>
<tr>
<td>Study selection</td>
<td>9</td>
<td>State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).</td>
<td>8-10</td>
</tr>
<tr>
<td>Data collection process</td>
<td>10</td>
<td>Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.</td>
<td>9-10</td>
</tr>
<tr>
<td>Data items</td>
<td>11</td>
<td>List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.</td>
<td>9-10</td>
</tr>
<tr>
<td>Risk of bias in individual studies</td>
<td>12</td>
<td>Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.</td>
<td>9-10</td>
</tr>
<tr>
<td>Summary measures</td>
<td>13</td>
<td>State the principal summary measures (e.g., risk ratio, difference in means).</td>
<td>10</td>
</tr>
<tr>
<td>Synthesis of results</td>
<td>14</td>
<td>Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$) for each meta-analysis.</td>
<td>10</td>
</tr>
<tr>
<td>Section/topic</td>
<td>#</td>
<td>Checklist item</td>
<td>Reported on page #</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----</td>
<td>---------------------------------------------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Risk of bias across studies</td>
<td>15</td>
<td>Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).</td>
<td>9-10</td>
</tr>
<tr>
<td>Additional analyses</td>
<td>16</td>
<td>Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.</td>
<td>NA</td>
</tr>
<tr>
<td><strong>RESULTS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study selection</td>
<td>17</td>
<td>Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.</td>
<td>10</td>
</tr>
<tr>
<td>Study characteristics</td>
<td>18</td>
<td>For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.</td>
<td>10</td>
</tr>
<tr>
<td>Risk of bias within studies</td>
<td>19</td>
<td>Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).</td>
<td>10</td>
</tr>
<tr>
<td>Results of individual studies</td>
<td>20</td>
<td>For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.</td>
<td>10-16</td>
</tr>
<tr>
<td>Synthesis of results</td>
<td>21</td>
<td>Present results of each meta-analysis done, including confidence intervals and measures of consistency.</td>
<td>10-16</td>
</tr>
<tr>
<td>Risk of bias across studies</td>
<td>22</td>
<td>Present results of any assessment of risk of bias across studies (see Item 15).</td>
<td>10-16</td>
</tr>
<tr>
<td>Additional analysis</td>
<td>23</td>
<td>Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).</td>
<td>NA</td>
</tr>
<tr>
<td><strong>DISCUSSION</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary of evidence</td>
<td>24</td>
<td>Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).</td>
<td>16-23</td>
</tr>
<tr>
<td>Limitations</td>
<td>25</td>
<td>Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).</td>
<td>23-24</td>
</tr>
<tr>
<td>Conclusions</td>
<td>26</td>
<td>Provide a general interpretation of the results in the context of other evidence, and implications for future research.</td>
<td>24</td>
</tr>
<tr>
<td><strong>FUNDING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funding</td>
<td>27</td>
<td>Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.</td>
<td>NA</td>
</tr>
</tbody>
</table>

For more information, visit: [www.prisma-statement.org](http://www.prisma-statement.org)
<table>
<thead>
<tr>
<th>Type of outcome measure</th>
<th>Outcomes measure</th>
<th>Reference no.</th>
<th>Age range</th>
<th>Mean Age</th>
<th>Gender (f%)</th>
<th>Ethnicity</th>
<th>Socioeconomic status</th>
<th>Country</th>
<th>Follow up sample size</th>
<th>Measure (F)</th>
<th>Effect size</th>
<th>Outcome monitoring system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brief Psychiatric Rating Scale (BPRS)</td>
<td>Follow up</td>
<td>17 - 30, 22</td>
<td>NA</td>
<td>Schizophrenia</td>
<td>Treatment effects</td>
<td>Community mental health service</td>
<td>NA</td>
<td>6 months &amp; 1 year</td>
<td>101</td>
<td>185%</td>
<td>NA</td>
<td>Italy</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>16 - 20, 23</td>
<td>NA</td>
<td>First episode psychosis</td>
<td>Randomised trial</td>
<td>Psychiatric teaching hospital</td>
<td>NA</td>
<td>3 &amp; 9 months</td>
<td>25</td>
<td>9%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>20 - 21, 19.9</td>
<td>NA</td>
<td>First episode psychosis</td>
<td>Treatment effects</td>
<td>Stabilisation for initial treatment and follow up</td>
<td>NA</td>
<td>2 &amp; 24 weeks</td>
<td>18</td>
<td>13%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>4 - 7, 18</td>
<td>NA</td>
<td>Schizophrenia</td>
<td>Treatment effects</td>
<td>Anxiety for acute and chronic schizophrenia</td>
<td>NA</td>
<td>Significance</td>
<td>23</td>
<td>20%</td>
<td>NA</td>
<td>England</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>16 - 20, 23.1</td>
<td>NA</td>
<td>Psychosis</td>
<td>Treatment effects</td>
<td>Individual and group treatment services</td>
<td>Treated = 6 - 90 weeks</td>
<td>1 year</td>
<td>193</td>
<td>0%</td>
<td>NA</td>
<td>America</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>4 - 22, 17.2</td>
<td>NA</td>
<td>Cognitive perseverative disorders</td>
<td>Treatment effects</td>
<td>Schizophrenia community</td>
<td>NA</td>
<td>Significance</td>
<td>11</td>
<td>10%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>15 - 20, 23.4</td>
<td>NA</td>
<td>First episode psychosis</td>
<td>Treatment effects</td>
<td>Early intervention and support</td>
<td>Significance</td>
<td>3 &amp; 12 months</td>
<td>25</td>
<td>2%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>1 - 4, 15.9</td>
<td>NA</td>
<td>Psychosis</td>
<td>Treatment effects</td>
<td>Outcome of the initial stage of mental health study in children and adolescents</td>
<td>Significance</td>
<td>1 week</td>
<td>25%</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>17 - 21, 24</td>
<td>NA</td>
<td>First episode psychosis</td>
<td>Treatment effects</td>
<td>Psychiatric treatment and rehabilitation center</td>
<td>Significance</td>
<td>6 months</td>
<td>41</td>
<td>40%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>16 - 20, 22.2</td>
<td>NA</td>
<td>Schizophrenia</td>
<td>Treatment effects</td>
<td>Centre for addiction and mental health</td>
<td>Significance</td>
<td>12 weeks</td>
<td>243</td>
<td>96%</td>
<td>NA</td>
<td>Canada</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>15 - 24, 19</td>
<td>NA</td>
<td>First episode psychosis</td>
<td>Treatment effects</td>
<td>Drug dependence assessment and prevention centre</td>
<td>Significance</td>
<td>6 weeks</td>
<td>120</td>
<td>9%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>12 - 14, 9.7</td>
<td>NA</td>
<td>Schizophrenia</td>
<td>Treatment effects</td>
<td>Early detection</td>
<td>Significance</td>
<td>Significance</td>
<td>6 &amp; 12 months</td>
<td>118</td>
<td>21%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>17 - 20, 15.9</td>
<td>NA</td>
<td>Mixed/Borderline</td>
<td>Treatment effects</td>
<td>Mental health centre</td>
<td>Significance</td>
<td>6 months</td>
<td>26</td>
<td>25%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>10 - 18, 14.7</td>
<td>Parents</td>
<td>Mixed/Borderline</td>
<td>Behaviour change</td>
<td>Psychiatric hospital</td>
<td>Significance</td>
<td>2 months</td>
<td>196</td>
<td>0%</td>
<td>NA</td>
<td>America</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 13.9</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Private psychiatric facility</td>
<td>Significance</td>
<td>Significance</td>
<td>6 months</td>
<td>172</td>
<td>48%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 12.2</td>
<td>NA</td>
<td>Mixed</td>
<td>Mixed randomised trial</td>
<td>Managed care organisation</td>
<td>Significance</td>
<td>6 months</td>
<td>279</td>
<td>42%</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>12 - 19, NA</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Multicultural service for at-risk and adolescents</td>
<td>Significance</td>
<td>6 months</td>
<td>20</td>
<td>13%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 17, 13.23</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Children's mental health agency</td>
<td>Significance</td>
<td>Significance</td>
<td>6 months</td>
<td>150</td>
<td>26%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 17, 13.23</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Community mental health centre</td>
<td>Significance</td>
<td>Significance</td>
<td>6 months</td>
<td>150</td>
<td>26%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>12 - 15.9</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Multicentre trial</td>
<td>Significance</td>
<td>6 months</td>
<td>105</td>
<td>3%</td>
<td>NA</td>
<td>America</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 17, 13.23</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Residential treatment centre</td>
<td>Significance</td>
<td>6 months</td>
<td>100</td>
<td>41%</td>
<td>NA</td>
<td>America</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>12 - 15, 15.2</td>
<td>Parents</td>
<td>Psychosis</td>
<td>Randomised trial</td>
<td>Centre for child and adolescent psychiatry</td>
<td>Significance</td>
<td>6 months</td>
<td>25</td>
<td>40%</td>
<td>NA</td>
<td>America</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 15.5</td>
<td>NA</td>
<td>Mixed</td>
<td>Mixed treatment effects</td>
<td>Residential treatment centre</td>
<td>Significance</td>
<td>6 months</td>
<td>1649</td>
<td>0%</td>
<td>NA</td>
<td>America</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>13 - 16, NA</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Community mental health centre</td>
<td>Treatment group CBCL Total = 17</td>
<td>Control CBCL, IN</td>
<td>12 weeks</td>
<td>79</td>
<td>95%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>20 - 14.9</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Mental health service</td>
<td>Significance</td>
<td>Significance</td>
<td>6 months</td>
<td>249</td>
<td>1%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 14</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>High welfare institution</td>
<td>Significance</td>
<td>Significance</td>
<td>6 months</td>
<td>430</td>
<td>2%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>12 - 13</td>
<td>Children</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Residential treatment centre</td>
<td>Significance</td>
<td>6, 12, 18, 24 &amp; 36 months</td>
<td>25</td>
<td>2%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>14 - 17, 15.7</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Child and youth mental health service</td>
<td>Effects sizes</td>
<td>6 months &amp; 3 months</td>
<td>5%</td>
<td>NA</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 17, 13.6</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Outpatient mental health centre</td>
<td>Significance</td>
<td>6 months</td>
<td>150</td>
<td>24%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 20, 16.2</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Mental health services</td>
<td>NA</td>
<td>2 years</td>
<td>359</td>
<td>37%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>12 - 15, 11</td>
<td>Parents</td>
<td>Mixed</td>
<td>Mixed randomised trial</td>
<td>Child and adolescent mental health services</td>
<td>Significance</td>
<td>6 months</td>
<td>53</td>
<td>20%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 12.3</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Outpatient mental health clinic</td>
<td>Significance</td>
<td>6 months</td>
<td>143</td>
<td>19%</td>
<td>NA</td>
<td>America</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>3 - 7, 12.7</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Mental health centre</td>
<td>Significance</td>
<td>Significance</td>
<td>12 months</td>
<td>51</td>
<td>8%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 17, 12.7</td>
<td>Parents</td>
<td>Mixed</td>
<td>Mixed treatment effects</td>
<td>Community mental health service</td>
<td>Significance</td>
<td>Significance</td>
<td>6 months</td>
<td>8</td>
<td>7%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 16.2</td>
<td>NA</td>
<td>Mixed</td>
<td>Mixed randomised trial</td>
<td>Community mental health services</td>
<td>Significance</td>
<td>Significance</td>
<td>6 months</td>
<td>82</td>
<td>0%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 17.6</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Psychiatric inpatient unit</td>
<td>Significance</td>
<td>2 weeks, 6, 9, 12 &amp; 18 months</td>
<td>12</td>
<td>36%</td>
<td>NA</td>
<td>America</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>12 - 15, 13.3</td>
<td>Parents</td>
<td>Depression</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Multicentre trial</td>
<td>Significance</td>
<td>6 months</td>
<td>192</td>
<td>44%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 14.2</td>
<td>Teacher</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Adolescent day treatment unit</td>
<td>Teacher - NA</td>
<td>2.5 &amp; 3.5 years</td>
<td>20</td>
<td>21%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 15.49</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Mental health clinic</td>
<td>Significance</td>
<td>5 weeks &amp; 3 months</td>
<td>80</td>
<td>15%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>12 - 15.32</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Centre for mental health services</td>
<td>Significance</td>
<td>Significance</td>
<td>6 months</td>
<td>122</td>
<td>1%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>12 - 15</td>
<td>Parents</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>University hospital</td>
<td>Significance</td>
<td>Significance</td>
<td>6 months</td>
<td>122</td>
<td>1%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 15.8</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Adolescent treatment for at-risk early adolescent</td>
<td>Significance</td>
<td>Significance</td>
<td>6 months</td>
<td>15</td>
<td>9%</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>12 - 18.14</td>
<td>Young female</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Transcultural mental health services</td>
<td>Significance</td>
<td>6 months</td>
<td>95</td>
<td>25%</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td>Follow up</td>
<td>11 - 20, 15.9</td>
<td>Young female</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Psychiatric inpatient unit</td>
<td>Significance</td>
<td>6 months</td>
<td>125</td>
<td>20%</td>
<td>NA</td>
<td>America</td>
</tr>
<tr>
<td>Study</td>
<td>Age</td>
<td>Setting</td>
<td>Intervention</td>
<td>Provider</td>
<td>Type</td>
<td>Duration</td>
<td>n</td>
<td>% Effective</td>
<td>Country</td>
<td>Notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>---------</td>
<td>--------------</td>
<td>----------</td>
<td>------</td>
<td>----------</td>
<td>---</td>
<td>-----------</td>
<td>---------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16 - 17, NA</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Service</td>
<td>Significance</td>
<td>1 year</td>
<td>104</td>
<td>46%</td>
<td>NA</td>
<td>England</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>16 - 15, NA</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Social welfare department</td>
<td>NA</td>
<td>Significance</td>
<td>1 year</td>
<td>199</td>
<td>65%</td>
<td>Chinese</td>
<td>England</td>
</tr>
<tr>
<td>3</td>
<td>16 - 15, NA</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>7 months &amp; 2 years</td>
<td>419</td>
<td>60%</td>
<td>English</td>
<td>England</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>16 - 17, NA</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Service</td>
<td>Significance</td>
<td>1 year</td>
<td>128</td>
<td>34%</td>
<td>NA</td>
<td>China</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>12 - 25, 19.2</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Social welfare department</td>
<td>NA</td>
<td>Significance</td>
<td>1 year</td>
<td>199</td>
<td>53%</td>
<td>Chinese</td>
<td>Australia</td>
</tr>
<tr>
<td>6</td>
<td>12 - 17, 17.4</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>9 months</td>
<td>168</td>
<td>58%</td>
<td>NA</td>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>17 - 25, 19.4</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>General community</td>
<td>Significance</td>
<td>1, 2, 3 &amp; 4 years</td>
<td>6151</td>
<td>NA</td>
<td>NA</td>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>16 - 16, 16</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>7 months &amp; 2 years</td>
<td>413</td>
<td>50%</td>
<td>English</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>12 - 25, 17.13</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Mixed</td>
<td>Significance</td>
<td>1 year</td>
<td>2937</td>
<td>54%</td>
<td>NA</td>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>12 - 10, NA</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>NA</td>
<td>Significance</td>
<td>1 year</td>
<td>31</td>
<td>52%</td>
<td>NA</td>
<td>NA</td>
<td>England</td>
</tr>
<tr>
<td>11</td>
<td>16 - 25, NA</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>3 &amp; 6 months</td>
<td>67</td>
<td>37%</td>
<td>NA</td>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>13 - 14, 13.38</td>
<td>Young person</td>
<td>None</td>
<td>Randomised trial</td>
<td>Schools</td>
<td>Effect size</td>
<td>9 weeks</td>
<td>36</td>
<td>56%</td>
<td>NA</td>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>17 - 24, NA</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>NA</td>
<td>Significance</td>
<td>1 year</td>
<td>2937</td>
<td>54%</td>
<td>NA</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td>14</td>
<td>12 - 18, NA</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Service</td>
<td>Significance</td>
<td>6 weeks &amp; 3 months</td>
<td>709</td>
<td>70%</td>
<td>NA</td>
<td>NA</td>
<td>England</td>
</tr>
<tr>
<td>15</td>
<td>12 - 16, 14.87</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>NA</td>
<td>Significance</td>
<td>2 &amp; 6 months</td>
<td>738</td>
<td>70%</td>
<td>NA</td>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>12 - 14, 14</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>NA</td>
<td>Significance</td>
<td>8 weeks &amp; 3 months</td>
<td>7</td>
<td>91%</td>
<td>NA</td>
<td>NA</td>
<td>England</td>
</tr>
<tr>
<td>17</td>
<td>12 - 18, 15</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>NA</td>
<td>Significance</td>
<td>Routine - 1 year</td>
<td>31</td>
<td>61%</td>
<td>NA</td>
<td>NA</td>
<td>England</td>
</tr>
<tr>
<td>18</td>
<td>11 - 19, 15</td>
<td>Parents</td>
<td>Young person</td>
<td>Mixed</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>Afghans</td>
<td>Primarily &quot;food insecure&quot; 64%</td>
<td>Afghanistan</td>
</tr>
<tr>
<td>19</td>
<td>12 - 18, 14.57</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>11 months and 1 year</td>
<td>41</td>
<td>82%</td>
<td>Primarily Swedish</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>20</td>
<td>12 - 16, 14</td>
<td>Parents</td>
<td>Young person</td>
<td>Mixed</td>
<td>Mixed</td>
<td>Significance</td>
<td>Routine - 16 weeks</td>
<td>356</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>21</td>
<td>11 - 16, 13</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>22</td>
<td>11 - 18, 14</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>23</td>
<td>11 - 16, 15.5</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>3 months</td>
<td>70%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>England</td>
</tr>
<tr>
<td>24</td>
<td>12 - 10, 15.5</td>
<td>NA</td>
<td>NA</td>
<td>Treatment</td>
<td>Child and adolescent mental health service</td>
<td>Significance</td>
<td>10.25 sessions &amp; 6 months after 10.25 sessions</td>
<td>23</td>
<td>100%</td>
<td>Chinese</td>
<td>Hong Kong</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>11 - 10, 15</td>
<td>Parents</td>
<td>Young person</td>
<td>Mixed</td>
<td>NA</td>
<td>Significance</td>
<td>1 year</td>
<td>31</td>
<td>31%</td>
<td>NA</td>
<td>NA</td>
<td>England</td>
</tr>
<tr>
<td>26</td>
<td>11 - 10, 15</td>
<td>Parents</td>
<td>Young person</td>
<td>Mixed</td>
<td>NA</td>
<td>Significance</td>
<td>1 year</td>
<td>31</td>
<td>31%</td>
<td>NA</td>
<td>NA</td>
<td>England</td>
</tr>
<tr>
<td>27</td>
<td>12 - 16, 14</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>NA</td>
<td>Significance</td>
<td>Routine - 5.5 sessions</td>
<td>202</td>
<td>57%</td>
<td>NA</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td>28</td>
<td>15 - 16, NA</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>NA</td>
<td>Significance</td>
<td>3 years</td>
<td>2489</td>
<td>55%</td>
<td>NA</td>
<td>NA</td>
<td>Norway</td>
</tr>
<tr>
<td>29</td>
<td>12 - 17, 14</td>
<td>Parents</td>
<td>Young person</td>
<td>Mixed</td>
<td>Schools</td>
<td>Significance</td>
<td>3 years</td>
<td>749</td>
<td>55%</td>
<td>NA</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td>30</td>
<td>11 - 16, 13</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>NA</td>
<td>Significance</td>
<td>Routine - 5.5 sessions</td>
<td>202</td>
<td>57%</td>
<td>NA</td>
<td>NA</td>
<td>Australia</td>
</tr>
<tr>
<td>31</td>
<td>12 - 18, 15.03, 14.57</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>11 months and 1 year</td>
<td>41</td>
<td>82%</td>
<td>Primarily Swedish</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>32</td>
<td>12 - 17, 12.5</td>
<td>Young person</td>
<td>None</td>
<td>Treatment</td>
<td>School</td>
<td>Significance</td>
<td>Routine - 5.5 sessions</td>
<td>33</td>
<td>33%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>33</td>
<td>12 - 18, 14</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>34</td>
<td>11 - 17, 14</td>
<td>Parents</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>Routine - 5.5 sessions</td>
<td>33</td>
<td>33%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>35</td>
<td>11 - 16, 15</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>36</td>
<td>11 - 16, 15</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>37</td>
<td>11 - 15, 15.5</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>38</td>
<td>11 - 16, 15</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>39</td>
<td>11 - 16, 15.5</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>40</td>
<td>11 - 15, 15</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>41</td>
<td>11 - 16, 15</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>42</td>
<td>11 - 16, 15</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>43</td>
<td>11 - 16, 15</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>44</td>
<td>11 - 16, 15</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Significance</td>
<td>1 year</td>
<td>234</td>
<td>50%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

**Effect sizes:**
- K-10 = .24, 9 weeks
- SDQ-emotional = .15, .35
- SDQ-impact = .51, 1.13
- Kessler Psychological Distress Scale (K10) = .24, 9 weeks
- Strengths and Difficulties Questionnaire (SDQ) = .15, .35
- Symptom and Functioning Severity Scale (SFSS) = .15, .35
- Symptom Checklist 90 Revised (SCL-90-R) = .15, .35
<table>
<thead>
<tr>
<th>Study ID</th>
<th>Age</th>
<th>Gender</th>
<th>Setting</th>
<th>Type of Assessment</th>
<th>Sample Size</th>
<th>Follow-up</th>
<th>Severity</th>
<th>Treatment Group</th>
<th>Effect Sizes</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1440</td>
<td>14 - 15, 15+</td>
<td>NA</td>
<td>Mixed</td>
<td>Randomised trial</td>
<td>Child and adolescent psychiatry department</td>
<td>102</td>
<td>8 weeks &amp; 3 months</td>
<td>ERT + TAU</td>
<td>Follow-up = 0.58</td>
<td>NA</td>
</tr>
<tr>
<td>1441</td>
<td>14 - 15, 15+</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Adolescent psychiatric unit</td>
<td>74</td>
<td>8 weeks &amp; 3 months</td>
<td>TAU</td>
<td>Primarily Caucasian 45%</td>
<td>NA</td>
</tr>
<tr>
<td>1442</td>
<td>NA, 15</td>
<td>Young person</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Private psychiatric facility</td>
<td>123</td>
<td>8 weeks &amp; 3 months</td>
<td>YSR subscale</td>
<td>Primarily Caucasian 51%</td>
<td>NA</td>
</tr>
<tr>
<td>1443</td>
<td>14 - 15, 15+</td>
<td>NA</td>
<td>Mixed</td>
<td>Randomised trial</td>
<td>Mental health centre</td>
<td>103</td>
<td>8 weeks &amp; 12 months</td>
<td>ERT + TAU</td>
<td>Follow-up = 0.31</td>
<td>NA</td>
</tr>
<tr>
<td>1444</td>
<td>11 - 14, 15+</td>
<td>NA</td>
<td>Young person</td>
<td>Mixed</td>
<td>Community mental health agency</td>
<td>95</td>
<td>12 - 24 months</td>
<td>YASR (18 - 30)</td>
<td>Primarily Caucasian 45%</td>
<td>NA</td>
</tr>
<tr>
<td>1445</td>
<td>11 - 17, 17+</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Community mental health service</td>
<td>90</td>
<td>6 months</td>
<td>YASR (18 - 30)</td>
<td>Primarily Caucasian 48%</td>
<td>NA</td>
</tr>
<tr>
<td>1446</td>
<td>12 - 25, 18+</td>
<td>NA</td>
<td>Young person</td>
<td>Mixed</td>
<td>Community mental health clinic</td>
<td>109</td>
<td>3 months</td>
<td>YSR subscale</td>
<td>Primarily Caucasian 51%</td>
<td>NA</td>
</tr>
<tr>
<td>1447</td>
<td>15 - 18, NA</td>
<td>NA</td>
<td>Borderline personality disorder</td>
<td>Non randomised trial</td>
<td>Mental health unit</td>
<td>110</td>
<td>12 months</td>
<td>TAU</td>
<td>Primarily Caucasian 51%</td>
<td>NA</td>
</tr>
<tr>
<td>1448</td>
<td>11 - 17, 17+</td>
<td>NA</td>
<td>Young person</td>
<td>Mixed</td>
<td>Children's mental health agency</td>
<td>95</td>
<td>12 months</td>
<td>TAU</td>
<td>Primarily Caucasian 51%</td>
<td>NA</td>
</tr>
<tr>
<td>1449</td>
<td>11 - 17, 17+</td>
<td>NA</td>
<td>Young person</td>
<td>Mixed</td>
<td>Community mental health centre</td>
<td>95</td>
<td>12 months</td>
<td>TAU</td>
<td>Primarily Caucasian 51%</td>
<td>NA</td>
</tr>
</tbody>
</table>

**TABLE 1:** Summary of different study designs and their outcomes. The table includes information on the age range, gender, setting, type of assessment, sample size, follow-up period, severity, treatment group, effect sizes, and significance of the results. The table is sorted by Study ID, with additional columns for Age, Gender, Setting, Type of Assessment, Sample Size, Follow-up, Severity, Treatment Group, Effect Sizes, and Significance.
<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Design</th>
<th>Treatment</th>
<th>Endpoint</th>
<th>N</th>
<th>Effect Size</th>
<th>Significance</th>
<th>Country/Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Residential: 7.8 months</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Residential mental health centres/home based service</td>
<td>6, 12, 20 &amp; 32 weeks</td>
<td>15-50%</td>
<td>Apache American Indian</td>
</tr>
<tr>
<td>2</td>
<td>Home based: 5.25 months</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Home based service</td>
<td>6, 9 &amp; 7 months</td>
<td>21-50%</td>
<td>NA, 16.2, N</td>
</tr>
<tr>
<td>3</td>
<td>Primarily multiethnic 21%</td>
<td>NA</td>
<td>NA</td>
<td>Treatment effects</td>
<td>Community mental health centres</td>
<td>6 months</td>
<td>21-50%</td>
<td>NA, 17.6, N</td>
</tr>
<tr>
<td>4</td>
<td>Primarily African American 76%</td>
<td>NA</td>
<td>NA</td>
<td>Treatment effects</td>
<td>Multi-site study</td>
<td>6 months</td>
<td>21-50%</td>
<td>NA, 17.8, N</td>
</tr>
<tr>
<td>5</td>
<td>Primarily Hispanic 52%</td>
<td>NA</td>
<td>NA</td>
<td>Treatment effects</td>
<td>Multi-site study</td>
<td>6 months</td>
<td>21-50%</td>
<td>NA, 17.9, N</td>
</tr>
<tr>
<td>6</td>
<td>Primarily Caucasian 72%</td>
<td>NA</td>
<td>NA</td>
<td>Treatment effects</td>
<td>Multi-site study</td>
<td>6 months</td>
<td>21-50%</td>
<td>NA, 17.10, N</td>
</tr>
</tbody>
</table>

**Notes:**
- N refers to the number of studies.
- Effect sizes are reported as either percentage or statistical significance.
- Significance levels are indicated as either 5% or 10%.
- Countries/regions listed include various geographical locations.
<table>
<thead>
<tr>
<th>Country</th>
<th>Age</th>
<th>Gender</th>
<th>Comparison</th>
<th>Design</th>
<th>Setting</th>
<th>Outcome</th>
<th>Significance</th>
<th>Effect Sizes</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium, Bulgaria, Czech Republic, Estonia, Germany, India, Italy, Poland, Romania, Russia, Spain, Ukraine, and America</td>
<td>Primarily Hispanic 71%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>43%</td>
<td>NA</td>
<td>America</td>
</tr>
<tr>
<td>Belgium, Bulgaria, Czech Republic, Estonia, Germany, India, Italy, Poland, Romania, Russia, Spain, Ukraine, and America</td>
<td>Primarily Caucasian 56%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>40%</td>
<td>NA</td>
<td>America</td>
</tr>
</tbody>
</table>

**Obsessive compulsive disorder**

<table>
<thead>
<tr>
<th>Country</th>
<th>Age</th>
<th>Gender</th>
<th>Comparison</th>
<th>Design</th>
<th>Setting</th>
<th>Outcome</th>
<th>Significance</th>
<th>Effect Sizes</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primarily Caucasian 73%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>79%</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Primarily Caucasian 70%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>70%</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

**Depression Treatment effects**

<table>
<thead>
<tr>
<th>Country</th>
<th>Age</th>
<th>Gender</th>
<th>Comparison</th>
<th>Design</th>
<th>Setting</th>
<th>Outcome</th>
<th>Significance</th>
<th>Effect Sizes</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primarily Caucasian 85%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>51%</td>
<td>NA</td>
<td>America</td>
</tr>
<tr>
<td>Primarily Caucasian 73%</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>73%</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Reference</td>
<td>(year)</td>
<td>NA</td>
<td>Treatment</td>
<td>Effect size</td>
<td>Treatment</td>
<td>Effect size</td>
<td>Follow-up</td>
<td>Effect size</td>
<td>Treatment</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>----</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>[20]</td>
<td>2015</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Community based clinic</td>
<td>214 days</td>
<td>29.4%</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[24]</td>
<td>2014</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Children's mental health agency</td>
<td>12.64 months</td>
<td>20.7%</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[25]</td>
<td>2013</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Government funded health service for young people</td>
<td>6, 12 &amp; 24 months</td>
<td>27.8%</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[26]</td>
<td>2012</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>School and psychiatric unit</td>
<td>6, 12, 18, 30, 42 &amp; 54 weeks</td>
<td>73.5%</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[27]</td>
<td>2011</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Child and adolescent mental health service</td>
<td>6, 12 &amp; 24 months</td>
<td>57%</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[28]</td>
<td>2010</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment effects</td>
<td>Mental health care institute</td>
<td>6, 12, 18 &amp; 24 months</td>
<td>35%</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

### Effect sizes:

- Mindfulness: $d = -1.04$ 1 year 101 33%
- GAF: $d = 3.34$ 10.81 days 75 51%
- EQ-5D: $d = 0.58$ 12 months 51 85%
- EQ-5D: $d = 1.17$ 1 month 17 85%
- EQ-5D: $d = -1.02$ 8 weeks & 3 months 74 73%
- GAI: $d = 1.01$ 1 week 11 93%
<table>
<thead>
<tr>
<th>Study</th>
<th>Age Group</th>
<th>Diagnosis</th>
<th>Setting</th>
<th>Treatment</th>
<th>Effect Sizes</th>
<th>Follow-Up</th>
<th>Significance</th>
<th>Country</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIDSCREEN-52</td>
<td>11 - 17, 14</td>
<td>Young person</td>
<td>Depression</td>
<td>Randomised trial</td>
<td>Mental health services, general population</td>
<td>EQ-5D d = 0.11</td>
<td>12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>KIDSCREEN-27</td>
<td>10 - 12, 8</td>
<td>Young person</td>
<td>Depression</td>
<td>Randomised trial</td>
<td>Mental health services, general population</td>
<td>EQ-5D d = 0.26</td>
<td>12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Medical Outcomes Study</td>
<td>10 - 18, 15</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>General community</td>
<td>Effect sizes: EORTC-QLQ-C30: TAU-Post = -0.14</td>
<td>6 &amp; 12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Mental Health Subscale</td>
<td>10 - 18, 15</td>
<td>Young person</td>
<td>None</td>
<td>Naturalistic</td>
<td>Schools</td>
<td>Effect sizes: EORTC-QLQ-C30: TAU-Post = -0.14</td>
<td>6 &amp; 12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 13, 12</td>
<td>Young person</td>
<td>Anxiety</td>
<td>Randomised trial</td>
<td>Mental health centres</td>
<td>Significance</td>
<td>6 &amp; 14 weeks</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 17, 15</td>
<td>Young person</td>
<td>Social anxiety</td>
<td>Randomised trial</td>
<td>Hospital</td>
<td>Significance</td>
<td>6 &amp; 12 weeks</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>16 - 18, 12</td>
<td>Young person</td>
<td>None</td>
<td>Randomised trial</td>
<td>School</td>
<td>Significance</td>
<td>6 weeks</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 13, 12</td>
<td>Young person</td>
<td>Anxiety</td>
<td>Randomised trial</td>
<td>Community cross-section study</td>
<td>Significance</td>
<td>6 &amp; 14 weeks</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 13, 12</td>
<td>Young person</td>
<td>Depression</td>
<td>Randomised trial</td>
<td>Mental health centres</td>
<td>Significance</td>
<td>6 &amp; 14 weeks</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 15, 16</td>
<td>Young person</td>
<td>Depression</td>
<td>Randomised trial</td>
<td>Private hospital</td>
<td>Significance</td>
<td>6 &amp; 12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 13, 12</td>
<td>Young person</td>
<td>Anxiety</td>
<td>Randomised trial</td>
<td>Private hospital</td>
<td>Significance</td>
<td>6 &amp; 12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 13, 12</td>
<td>Young person</td>
<td>Anxiety</td>
<td>Randomised trial</td>
<td>Private hospital</td>
<td>Significance</td>
<td>6 &amp; 12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 13, 12</td>
<td>Young person</td>
<td>Anxiety</td>
<td>Randomised trial</td>
<td>Private hospital</td>
<td>Significance</td>
<td>6 &amp; 12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 13, 12</td>
<td>Young person</td>
<td>Anxiety</td>
<td>Randomised trial</td>
<td>Private hospital</td>
<td>Significance</td>
<td>6 &amp; 12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 13, 12</td>
<td>Young person</td>
<td>Anxiety</td>
<td>Randomised trial</td>
<td>Private hospital</td>
<td>Significance</td>
<td>6 &amp; 12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 13, 12</td>
<td>Young person</td>
<td>Anxiety</td>
<td>Randomised trial</td>
<td>Private hospital</td>
<td>Significance</td>
<td>6 &amp; 12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
<tr>
<td>Health of the Nation Outcome Scale</td>
<td>10 - 13, 12</td>
<td>Young person</td>
<td>Anxiety</td>
<td>Randomised trial</td>
<td>Private hospital</td>
<td>Significance</td>
<td>6 &amp; 12 months</td>
<td>Primarily Caucasian 76%</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note: The table above represents a sample of data from various studies focusing on different aspects of mental health outcomes. Each row provides information on the study name, age group, diagnosis, setting, treatment, effect sizes, follow-up, significance, country, and region. The data is organized in a tabular format for easy reading and analysis.
<table>
<thead>
<tr>
<th>Source</th>
<th>Study Duration</th>
<th>Setting</th>
<th>Intervention</th>
<th>Effect</th>
<th>Effect Size</th>
<th>Significance</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>[245]</td>
<td>4 - 17, 18.42</td>
<td>NA</td>
<td>Depression</td>
<td>Treatment</td>
<td>Effects</td>
<td>Mental health institutions</td>
<td>Omega squared = .04</td>
</tr>
<tr>
<td>[246]</td>
<td>14 - 17, NA</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Mental health service</td>
<td>Significance</td>
</tr>
<tr>
<td>[247]</td>
<td>13 - 17, 15.2</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>University staff and additional staff practice</td>
<td>Significance</td>
</tr>
<tr>
<td>[248]</td>
<td>14 - 17, 16.4</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Medical and psychological unit</td>
<td>Significance</td>
</tr>
<tr>
<td>[249]</td>
<td>13 - 17, 15.1</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Child and adolescent mental health service</td>
<td>Significance</td>
</tr>
<tr>
<td>[250]</td>
<td>2 - 15, 15.4</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>10 patient tools for adolescents</td>
<td>Significance</td>
</tr>
<tr>
<td>[251]</td>
<td>11 - 17, 14</td>
<td>Young person</td>
<td>Anorexia nervosa</td>
<td>Randomised trial</td>
<td>Treatment</td>
<td>Non-scheduled trial</td>
<td>Significance</td>
</tr>
<tr>
<td>[252]</td>
<td>13 - 18, 15.05</td>
<td>Young person</td>
<td>Anorexia nervosa</td>
<td>Randomised trial</td>
<td>Treatment</td>
<td>Non-scheduled trial</td>
<td>Significance</td>
</tr>
<tr>
<td>[253]</td>
<td>11 - 17, 14</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>University child and adolescent psychiatry service</td>
<td>Significance</td>
</tr>
<tr>
<td>[254]</td>
<td>12 - 18, 15.1</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Adolescent mental health service</td>
<td>Significance</td>
</tr>
<tr>
<td>[255]</td>
<td>12 - 17, 14</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Adolescent inpatient hospital unit</td>
<td>Significance</td>
</tr>
<tr>
<td>[256]</td>
<td>12 - 17, 14</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Child and adolescent mental health service</td>
<td>Significance</td>
</tr>
<tr>
<td>[257]</td>
<td>11 - 18, NA</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Child and adolescent mental health service</td>
<td>Significance</td>
</tr>
<tr>
<td>[258]</td>
<td>11 - 17, 14</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Adolescent inpatient hospital unit</td>
<td>Significance</td>
</tr>
<tr>
<td>[259]</td>
<td>13 - 18, 15.05</td>
<td>Young person</td>
<td>Anorexia nervosa</td>
<td>Randomised trial</td>
<td>Treatment</td>
<td>Non-scheduled trial</td>
<td>Significance</td>
</tr>
<tr>
<td>[260]</td>
<td>12 - 17, 14</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>In-patient facility for adolescents</td>
<td>Significance</td>
</tr>
<tr>
<td>[261]</td>
<td>11 - 17, 14</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Adolescent mental health service</td>
<td>Significance</td>
</tr>
<tr>
<td>[262]</td>
<td>12 - 18, 15.1</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Adolescent mental health service</td>
<td>Significance</td>
</tr>
<tr>
<td>[263]</td>
<td>11 - 17, 14</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Adolescent mental health service</td>
<td>Significance</td>
</tr>
<tr>
<td>[264]</td>
<td>11 - 17, 14</td>
<td>Clinician</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Adolescent mental health service</td>
<td>Significance</td>
</tr>
</tbody>
</table>

**Ohio Scales**

<table>
<thead>
<tr>
<th>Source</th>
<th>Study Duration</th>
<th>Setting</th>
<th>Intervention</th>
<th>Effect</th>
<th>Effect Size</th>
<th>Significance</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>[265]</td>
<td>4 - 10, 12.2</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Non-randomised trial</td>
<td>Significance</td>
</tr>
<tr>
<td>[266]</td>
<td>12 - 18, 14.20</td>
<td>Parent</td>
<td>Young person</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Community mental health centre</td>
<td>Significance</td>
</tr>
<tr>
<td>[267]</td>
<td>NA, 12.54</td>
<td>NA</td>
<td>Young person</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Community mental health centre</td>
<td>Significance</td>
</tr>
<tr>
<td>[268]</td>
<td>12 - 10, 18.82</td>
<td>Parent</td>
<td>Young person</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Department of mental health</td>
<td>Significance</td>
</tr>
<tr>
<td>[269]</td>
<td>12 - 10, NA</td>
<td>Young person</td>
<td>Mixed</td>
<td>Treatment</td>
<td>School counselling</td>
<td>Significance</td>
<td>6 months</td>
</tr>
<tr>
<td>[270]</td>
<td>12 - 14.2</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>Community mental health centres</td>
<td>Significance</td>
</tr>
</tbody>
</table>

**Young Persons Clinical Outcomes for Routine Evaluation (YPE CORE)**

<table>
<thead>
<tr>
<th>Source</th>
<th>Study Duration</th>
<th>Setting</th>
<th>Intervention</th>
<th>Effect</th>
<th>Effect Size</th>
<th>Significance</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>[271]</td>
<td>6 - 15, 13.7</td>
<td>Parent</td>
<td>Young person</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Community mental health and private managed care</td>
<td>Significance</td>
</tr>
<tr>
<td>[272]</td>
<td>6 - 13, 16.3</td>
<td>NA</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Effects</td>
<td>School</td>
<td>Significance</td>
</tr>
</tbody>
</table>

**Youth Outcome Questionnaire (Y-OQ)**

<table>
<thead>
<tr>
<th>Source</th>
<th>Study Duration</th>
<th>Setting</th>
<th>Intervention</th>
<th>Effect</th>
<th>Effect Size</th>
<th>Significance</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>[273]</td>
<td>4 - 17, 12.2</td>
<td>Parent</td>
<td>Young person</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Community mental health and private managed care</td>
<td>Significance</td>
</tr>
<tr>
<td>[274]</td>
<td>12 - 17, 12.3</td>
<td>Parent</td>
<td>Young person</td>
<td>Mixed</td>
<td>Treatment</td>
<td>Community mental health and private managed care</td>
<td>Significance</td>
</tr>
<tr>
<td>[275]</td>
<td>4 - 16, 16</td>
<td>Young person</td>
<td>None</td>
<td>Treatment</td>
<td>Effects</td>
<td>Schools</td>
<td>Significance</td>
</tr>
</tbody>
</table>
3 November 2015

University of Canberra Committee for Ethics in Human Research

Dear Chair

This letter is to confirm that headspace National Office (hNO) agrees to provide a timeshot of headspace centre data for use in a Psychology Clinical PhD project to be undertaken in 2015/16 by Mr Ben Kwan under the supervision of Professor Debra Rickwood.

The project is examining the reliability and validity of the MyLifeTracker measure as an outcome measure for headspace clients.

The timeshot is of data collected from January 2015 to June 2015. The data have been stripped of all identifying information and are non-identifiable. The headspace data are collected primarily for the purposes of service quality, monitoring and reporting, and clients consent to their information being used for research and evaluation projects that are related to service improvement. This current project will provide information about better ways to measure outcomes in early intervention for the mental health of young people.

The data timeshot must be stored securely on Professor Rickwood's computer and accessed only by Professor Rickwood and Mr Kwan. The data are not to be used for any other purpose and headspace requests a copy of the final thesis. Publications and presentations from the thesis are permissible, provided these are first checked by headspace National Office to ensure they are consistent with the agreed aims.

Yours sincerely

Kathleen Alonso
Chief Operating Officer
30 November 2015

Mr Benjamin Kwan
Faculty of Health
University of Canberra
Canberra ACT 2601

Dear Benjamin,

The Human Research Ethics Committee has considered your application to conduct research with human subjects for the project titled **MyLife Tracker: A pilot study of a routine outcome measure used with young people aged 12 to 25 years.**

**Approval is granted until 31 December 2016.**

The following general conditions apply to your approval.

These requirements are determined by University policy and the *National Statement on Ethical Conduct in Human Research* (National Health and Medical Research Council, 2007).

<table>
<thead>
<tr>
<th>Monitoring:</th>
<th>You must, in conjunction with your supervisor, assist the Committee to monitor the conduct of approved research by completing and promptly returning project review forms, which will be sent to you at the end of your project and, in the case of extended research, at least annually during the approval period.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discontinuation of research:</td>
<td>You must, in conjunction with your supervisor, inform the Committee, giving reasons, if the research is not conducted or is discontinued before the expected date of completion.</td>
</tr>
<tr>
<td>Extension of approval:</td>
<td>If your project will not be complete by the expiry date stated above, you must apply in writing for extension of approval. Application should be made before current approval expires; should specify a new completion date; should include reasons for your request.</td>
</tr>
<tr>
<td>Retention and storage of data:</td>
<td>University policy states that all research data must be stored securely, on University premises, for a minimum of five years. You must ensure that all records are transferred to the University when the project is complete.</td>
</tr>
<tr>
<td>Contact details and notification of changes:</td>
<td>All email contact should use the UC email address. You should advise the Committee of any change of address during or soon after the approval period including, if appropriate, email address(es).</td>
</tr>
</tbody>
</table>

Yours sincerely
Human Research Ethics Committee

---

Hendryk Flaegel
Research Integrity & Ethics
Research Services
T (02) 62420150 F (02) 62420151 E hendryk.flaege@canberra.edu.au
Dear Ben

The Human Research Ethics Committee has considered your application to conduct research with human subjects for the project "0468 - MyLifeTracker within a measurement feedback system".

The Committee made the following evaluation: Approved

The approval is valid until: 31/12/2019

The following general conditions apply to your approval. These requirements are determined by University policy and the National Statement on Ethical Conduct in Human Research (National Health and Medical Research Council, 2007).

**Monitoring**

You must assist the Committee to monitor the conduct of approved research by completing project review forms, and in the case of extended research, at least annually during the approval period.

**Reporting Adverse Events**

You must report any unexpected adverse events or complications that occur anytime during the conduct of the research study or during the follow up period after the research. Please refer these matters promptly to the HREC. Failure to do so may result in the withdrawal of the Ethics approval.

**Discontinuation of Research**

You must inform the Committee, giving reasons, if the research is not conducted or is discontinued before the expected date of completion.

**Extension of Approval**

If your project will not be complete by the expiry date stated above, you must apply for extension of approval. This must be done before current approval expires.

**Retention and Storage of Data**

University policy states that all research data must be stored securely, on University premises, for a minimum of five years. You must ensure that all records are transferred to the University when the project is complete.

**Contact Details and Notification of Changes**

All email contact should use the UC email address. You should advise the Committee of any change of address during or soon after the approval period including, if appropriate, email address(es).
Please do not hesitate to contact us via email if you require any further information.

All the best,

Hendryk Flaegel

Research Ethics & Integrity

Research Services

University of Canberra

15 March 2019
Invitation to headspace clinicians to participate in research

We are looking to recruit clinicians (psychiatrists, GPs, psychologists, intern psychologists, social workers, youth workers, occupational therapists, nurses, and other practitioners) from headspace centres across Australia who have access to the client outcome measures collected through the hAPI system.

Participation will involve completing a survey about how MyLifeTracker is used as a measurement feedback system in youth mental health settings. The survey will take approximately 10 minutes to complete and results will be entered into a secure database for analysis.

Participation is voluntary and participants are able to withdraw at any time without penalty. All information collected will remain anonymous and no identifying information is collected.

For more information please click on the link below that will take you to the participant information form.

http://canberrahealth.az1.qualtrics.com/jfe/form/SV_a9nEMPPJVNtqz6B

The survey will be open until the 24th April 2019.

Principal Researcher: Ben Kwan, PhD Candidate, Faculty of Health, University of Canberra;
Supervisor: Professor Debra Rickwood, Faculty of Health, University of Canberra & Chief Scientific Advisor, headspace National.
Clinician Participant Information Form

**Project title:** The use of MyLifeTracker as a measurement feedback system in youth mental health settings

**Researchers:**
Principal Researcher: Ben Kwan, Doctor of Philosophy in Clinical Psychology, Faculty of Health, University of Canberra;
Supervisor: Professor Debra Rickwood, Faculty of Health, University of Canberra.

**Project Aim:**
The aim of this study is to explore clinicians’ views on the use of MyLifeTracker as a measurement feedback system in youth mental health settings. MyLifeTracker is a brief mental health outcome measure designed for young people aged 12–25 years. Its items measure current self-reported quality of life in five different areas of importance to young people: general well-being, day-to-day activities, relationships with friends, relationships with family, and general coping. The measure is intended to be used at every session. It is completed electronically on a tablet or desktop device, and the information is immediately available to the clinician in the visual representation of graphs over time (via the hAPI system). MyLifeTracker has been tested to be psychometrically sound for young people and shows evidence of a single factor structure.

**Participant Involvement:**
We are looking to recruit a range of clinicians e.g. psychologists, intern psychologists, GPs, psychiatrists, social workers, youth workers, occupational therapists, nurses, and other practitioners from headspace centres across Australia that have access to MyLifeTracker in their work. Participation will involve completing a survey about the use of MyLifeTracker as a measurement feedback system in youth mental health settings. Results will inform a research project to be completed as a course requirement for the Doctor of Philosophy in Clinical Psychology program as well as provide headspace with information about the utility of MyLifeTracker in clinical practice.

The survey will take approximately 10 minutes to complete and results will be entered into a secure database for analysis. Participation in this study is voluntary and participants are able to withdraw at any time without penalty. It is not anticipated that participation in this study will present any risk to participants.
Confidentiality, Anonymity and Data Storage:

All information collected from the survey will remain anonymous and no identifying information will be collected. Only the researchers will have access to the data collected and all responses will be treated with complete confidentiality. All data will be secured, stored and disposed of according to University guidelines (stored at the University for 5 years).

Ethics Committee Approval:

This study has been approved by the University of Canberra Committee for Ethics in Human Research, Project Number 0468. The Research Ethics & Integrity team can be contacted by email humanethicscommittee@canberra.edu.au.

Further queries:

A summary of the results will be made available through headspace after completion. For any further inquiries or information please contact the principal researcher Ben Kwan ( ). The research is being supervised by Professor Debra Rickwood ( ). Please feel free to contact her regarding any concerns you may have about the research.

If you consent to participate in the online survey please tick the circle below and proceed to the next page.

☐ I consent to participate in this study. I have read and understood the purpose and requirements of this study.
Demographics

Age: (Free text box)

Gender: (Male, Female, Other)

Profession: (Psychologist, Social worker, GP, Psychiatrist, Nurse, Youth Worker, Drug and Alcohol Worker, Mental Health Worker, Counsellor, Other)

Years of clinical mental health experience (including external to headspace): (Free text box)

Highest completed education level: (High School, Certificate, Diploma, Bachelor’s degree, Master’s degree, Doctoral degree)

headspace centre: (Free text box)

How many years have you worked for headspace? (Free text box)

Main funding source: (Private practitioner, In-kind partner, On-staff grant funded, Other)

Primary therapeutic stance: (CBT, Psychodynamic/psychoanalytic, Family systems, Humanistic/client centred, Eclectic, Other orientation)

Are you a clinician with headspace’s Youth Early Psychosis Program (hYEPP)? (Yes, No)

Do you currently provide counselling/therapy at headspace: (Yes, No)
Monitoring and Feedback Attitudes Scale (MFA)

These items ask about routine progress monitoring and providing feedback to clients about treatment progress:

“Routine progress monitoring” refers to administering measures to your clients every 1-2 sessions in order to monitor their progress in treatment.

Providing “feedback” refers to discussing data that have been gathered through routine progress monitoring with clients (for example, a clinician might show a client a graph of their scores from previous sessions to facilitate a discussion of why treatment is or is not progressing as intended or what has led to treatment success).

Please use the following scale to rate how much you agree with the statements below:

Strongly Disagree (1) Disagree (2) Neutral (3) Agree (4) Strongly Agree (5)

- Monitoring treatment progress is an important part of treatment
- Monitoring treatment progress is valuable for supervision
- Providing feedback to clients about treatment progress helps to increase client motivation and engagement
- Providing clients with feedback about their treatment progress empowers them to make informed decisions about their care
- Providing clients with feedback about treatment progress facilitates collaboration between clients and clinicians
- Clients want their therapists to provide them with information about treatment progress
- Providing clients with feedback about treatment progress can increase their insight
- Providing clients with feedback about treatment progress helps keep treatment focused on treatment goals
- Providing clients with regular feedback about treatment progress creates an expectation for positive change
- Providing feedback to clients about treatment progress (or lack thereof) can lead to better treatment outcomes

- Providing feedback to clients about treatment progress (or lack thereof) would potentially harm the therapeutic alliance
- Providing clients with negative feedback about their progress would lead to client deterioration or premature treatment termination
- Providing clients with negative feedback about their progress would decrease their motivation for and/or engagement in treatment
- Providing clients with negative feedback about their progress would make them think their therapist is incompetent
MyLifeTracker measure and graph examples

A brief mental health measure called MyLifeTracker (MLT) has been developed for use in headspace as a “routine progress monitor”. For your reference for the subsequent questions, the measure is shown below, along with the types of graphs of change over time that are produced. Please have a good look at the measure and the graphs and then click next to continue to the questions.
Adapted from Attitudes Toward Standardized Assessment Scales-Monitoring and Feedback (ASA-MF)

The following questions ask about use of MyLifeTracker for “routine progress monitoring”. This means using the measure with clients every 1-2 sessions to monitor their treatment progress.

Providing “feedback” refers to discussing data that have been gathered through routine progress monitoring with clients (for example, a clinician might show a client a graph of their scores from previous sessions to facilitate a discussion of why treatment is or is not progressing as intended or what has led to treatment success).

Please answer the questions based on MyLifeTracker specifically, not on your experience of the entire hAPI system.

Please use the following scale to rate how much you agree with the statements below:

Strongly Disagree (1) Disagree (2) Neutral (3) Agree (4) Strongly Agree (5)

- MyLifeTracker doesn’t tell me anything I can’t learn from just talking to clients
- Using clinical judgment to monitor progress is superior to using MyLifeTracker
- MyLifeTracker provides more useful information than other assessments like informal interviews or observations
- MyLifeTracker doesn’t capture what’s really going on with clients
- Clinical problems are too complex to be captured by MyLifeTracker
- MyLifeTracker gathers information about the client that may not otherwise come up in session
- MyLifeTracker is not able to detect meaningful changes as they occur
- MyLifeTracker doesn’t measure the outcome domains most important to clients

- MyLifeTracker helps gather objective information about whether treatment is working
- MyLifeTracker helps identify when treatment is not going well
- MyLifeTracker can provide helpful information about whether it is time to terminate treatment
- Information from MyLifeTracker can help me plan for sessions
- MyLifeTracker helps identify when to change the overall treatment plan

- MyLifeTracker can efficiently gather information
- The information I receive from MyLifeTracker isn’t worth the time spent accessing, reviewing and interpreting the results
- MyLifeTracker interferes with establishing rapport during a session
- Completing MyLifeTracker is too much of a burden for my clients
- I do not have time to access and review MyLifeTracker scores on a frequent basis
Organisational Factors for MyLifeTracker Use scale

Please use the following scale to rate how much you agree with the statements below:

Strongly Disagree (1) Disagree (2) Neutral (3) Agree (4) Strongly Agree (5)

- The workplace has implemented MyLifeTracker into an electronic system that has easy accessibility to information
- The workplace has implemented MyLifeTracker into an electronic system that has a good design and layout
- I have adequate time allocated to me to use MyLifeTracker for routine progress monitoring and providing feedback
- I have adequate technological resources to use MyLifeTracker for routine progress monitoring and providing feedback (e.g. computer or tablet devices or printers)
- I have adequate technical support to use MyLifeTracker for routine progress monitoring and providing feedback (e.g. IT support & support using measurement system for any technical faults and errors)
- I have adequate administrative support for routinely administrating MyLifeTracker (e.g. reception and administration workers)
- The organisation encourages me to use MyLifeTracker with treatment planning and providing feedback within sessions
- Work colleagues encourage me to use MyLifeTracker with treatment planning and providing feedback within sessions
- My supervisor/manager encourages me to use MyLifeTracker with treatment planning and providing feedback within sessions
- The workplace policy dictates my use of MyLifeTracker with treatment planning and providing feedback within sessions
- I see colleagues using MyLifeTracker to inform treatment planning and providing feedback within sessions
- Colleagues don’t find MyLifeTracker useful
- The team are generally positive about using MyLifeTracker
- I was provided an adequate induction on MyLifeTracker when I first started at headspace/my workplace
- My supervisor/manager provides clinical support of MyLifeTracker with treatment planning and providing feedback within sessions
- MyLifeTracker scores are discussed in clinical review meetings as a part of treatment planning for a client
- I receive ongoing clinical support from my organisation around using MyLifeTracker
- I have adequate training with using MyLifeTracker in treatment planning
- I have adequate training about how to provide feedback of MyLifeTracker scores to clients
- I have adequate training on how to interpret the results of MyLifeTracker
- I have access to training on MyLifeTracker
MyLifeTracker Monitoring and Feedback Use

These items ask about how often you use MyLifeTracker.

- What percentage of sessions do you look at MyLifeTracker before session?
- What percentage of sessions do you look at MyLifeTracker after session?
- What percentage of sessions do use MyLifeTracker to inform treatment planning in any way (i.e. alter plan, continue plan as it supports it, add more supports)?
- What percentage of sessions do you provide feedback of MyLifeTracker scores to clients in any way (i.e. discuss results, show graph results, ask about areas of change/deterioration or no change)?

MyLifeTracker Monitoring and Feedback Use Qualitative Questions

These items ask about factors that could help allow you to use MyLifeTracker more often.

- Which sessions or under what conditions do you look at MyLifeTracker before session?
- Which sessions or under what conditions do you look at MyLifeTracker after session?
- Which sessions or under what conditions do you use MyLifeTracker to inform treatment planning?
- Which sessions or under what conditions do you provide feedback of MyLifeTracker to clients?

- What do you think would help allow you to look at MyLifeTracker before session more often?
- What do you think would help allow you to look at MyLifeTracker after session more often?
- What do you think would help allow you to use MyLifeTracker to inform treatment planning more often?
- What do you think would help allow you to provide feedback of MyLifeTracker scores to clients more often (i.e. jointly discuss progress, review outcome, and make changes to the treatment plan if needed)?

Thank you for your time spent taking this survey. Much appreciated. Your responses have been recorded. A summary of the results will be available via ourspace and through the centre bulletin later in the year.